现代应用数学丛书

数值计算法

〔日〕森口繁一　高田　胜　著

上海科学技才

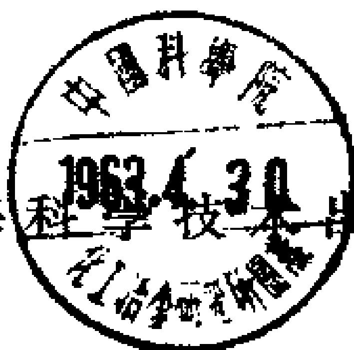
现代应用数学丛书

数值計算法

〔日〕森口繁一 高田 勝 著

閻昌齡 譯

張鴻 游兆永 校



上海科学技术出版社

內 容 提 要

本书是日本岩波书店出版的現代应用数学丛书之一的中譯本。全书共分 7 章，討方程的解法，差分与插补，数値积分·数値微分，函数逼近，常微分方程、偏微分方程和特征値問題的数値解法。适宜于初学者以及从事于实际数値計算工作的数学工作者或工程师参考，也可供高等院校作为教学参考书。

現代应用数学丛书

数 値 計 算 法

原 书 名　数 値 計 算 法

原 著 者　〔日〕森口繁一 高田勝

原出版者　岩波书店 1958

譯　　者　閻 昌 龄

校　　者　張 鴻 游兆永

*

上海科学技术出版社出版

（上海瑞金二路 450 号）

上海市书刊出版业营业許可証出 093 号

新华书店上海发行所发行　各地新华书店經售

商务印书館上海厂印刷

*

开本 850×1168 1/32　印張 6 2/32　字数 141,000

1963 年 3 月第 1 版　1963 年 3 月第 1 次印刷

印数 1—6,000

統一书号：13119·501

定　　价：（十四）1.05 元

出版說明

这一套书是根据日本岩波书店出版的"現代应用数学讲座"翻譯而成。日文原书共15卷60册，分成A、B两組，各編有序号。現在把原来同一题目分成两册或三册的加以合并，整理成42种，不另分組編号，陆續翻譯出版。

这套书涉及的面很广，其內容都和現代科学技术密切有关，有一定参考价值。每一本书收集的資料都比較丰富，而叙述扼要，篇幅不多，有利于讀者以較短时間掌握有关学科的主要內容。虽然，这套书的某些观点不尽适合于我国的情况，但其方法可供参考。因此，翻譯出版这一套书，对我国学术界是有所助益的。

由于日文原书是1957年起以讲座形式陆續出版的，写作时間和篇幅的限制不可避免地会影响原作者对內容的处理，为了尽可能地減少这种影响，我們在每一譯本中，特請譯者或校閱者撰写序或后記，以介紹有关学科的最近发展状况，并对全书內容作一些評价，提出一些看法，結合我国情况补充一些資料文献，在文內过于簡略或不足的地方添加了必要的注释和改正原书中存在的一些錯誤。希望这些工作能对讀者有所帮助。

承担翻譯和校閱的同志，为提高书籍的质量付出了巨大劳动，在此特致以誠摯的謝意。

欢迎讀者对本书提出批評和意見。

上海科学技术出版社

现代应用数学丛书

书名	原作者	譯者
代数学*	弥永昌吉等	熊全淹
几何学*	矢野健太郎	孙澤瀛
复变函数	功力金二郎	刘书琴
集合·拓扑·测度*	河田敬义	賴英华
泛函分析*	吉田耕作	程其襄
广义函数*	岩村联	楊永芳
常微分方程*	福原滿洲雄	張庆芳
偏微分方程*	南云道夫	錢端壮
特殊函数*	小谷正雄等	錢端壮
差分方程*	福田武雄	穆鴻基
富里哀变换与拉普拉斯变换*	河田龙夫	錢端壮
变分法及其应用*	加藤敏夫	周怀生
李群論*	岩堀長庆	孙澤瀛
随机过程*	伊藤清	刘璋温
回轉群与对称群的应用*	山内恭彦等	張质賢
結晶統計与代数*	伏見康治	孙澤瀛
偏微分方程的应用	犬井鉄郎等	楊永芳
微分方程的近似解法	加藤敏夫等	王占瀛
数值計算法	森口繁一等、	閻昌齡
量子力学中的数学方法*	朝永振一郎	周民强
工程力学系統*	近藤一夫等	刘亦珩
非綫性振动論*	古屋茂	吕紹明
力学系与映射理論*	岩田义一	孙澤瀛
平面彈性論*	森口繁一	刘亦珩
有限变位彈性論、变形几何学	山本善之、近藤一夫	刘亦珩
塑性論*	鷲津文一郎	刘亦珩
粘性流体理論*	谷一郎	刘亦珩
可压縮流体理論*	河村龙馬	刘亦珩
网絡理論	喜安善市等	賈弄晳
自动控制理論*	喜安善市等	瞿立林
回路拓扑学	近藤一夫	張鳴鏞
信息論*	喜安善市等	李文清
推断統計理論	北川敏男	李賢平
統計分析*	森口繁一	刘璋温
試驗設計法	增山元三郎	刘璋温
群体遺傳学的数学理論*	木村資生	刘祖洞
博奕論	宫澤光一	張鍾春
綫性規划	森口繁一	刘源張
經济理論中的数学方法	安井琢磨等	談祥柏
随机过程的应用*	河田龙夫	刘璋温
計算技术	高橋秀俊	姚晋
穿孔卡計算机	森口繁一	刘源張

注：有 * 者已在 1962 年出版。

序

数值計算法，又有数值分析、实用分析等等的名称。这門学科虽然有着与其說是科学无宁說是技术更为恰当的一面，但因为最近以来不論在理論方面还是在实际方面都取得了显著的发展，因此，事实上已經成为(数学的)一个卓越的分支。

現代并用着从笔算一直到台式計算机的“手工計算”和从穿孔卡計算机一直到快速电子計算机的“机械計算”。由于各种計算手段按照它在性能上和經济上的不同特征各有其适用范圍，因此这种状况今后还会继續存在下去。不过，总的說来，在今后的一段时期內机械計算所占的比重将要有急驟的增长，而且在手工計算方面，計算机所占的比重无疑也还要进一步增长。面对着这样的形势，出現数值計算法的飞跃发展是可以預期的。

本书的目的在于把数值計算中常用的方法以及这些方法中的难点和处理法作一个概括的介紹。内容并不全面，而是尽可能地作了精选，并力求使它容易看懂。例題是考虑到使用 10 位 ×10 位＝20 位左右的台式計算机而选取的，但这些例題的精神实质，大多数对于位数更少或更多的情形，乃至对于快速电子計算机也是适用的。这一类的学习只凭看书是不行的，自己編造类似的例題，并利用附近可以利用到的計算工具实际练习着計算是最重要的。这是作者們实行过的学习方法，并且愿意恳切地把它推荐給讀者。

希望本书能有助于理工方面以及其他方面的研究工作者以更高的效率进行在他們工作中所必要的計算。同时也非常希望爱好数学的广大讀者能够由于发生兴趣而进入这一分支，努力于計算

的实施并进一步发展有关的理論和方法，如果本书能成为这样的推动力，作者是很欣慰的。

第一章到第四章由森口执笔，第五章以后是高田写的而由森口作了修改。此外，在計算以及其他方面得到了伊东繁的协助，志此以表示謝意。

森　口

高　田　1957年11月23日

目　录

第1章 方程的解法

§1 綫性計算(扫除法)

联立1次方程(例如后面的(1.7))的解法是自古以来就被研究着的問題。作为理論上的解法有著名的 Cramer 公式。这公式把解表示成分数,以系数的行列式为分母,而以方程的右边代换相应的列后所得的行列式为分子。但是,在n元的情形应用这一公式,假定要按定义計算其中的$n+1$个行列式的值,因为計算每个行列式的$n!$个項的每一項都需要作$n-1$次乘法,因此总共就需要作$n!(n-1)(n+1)$次乘法。(此外还有加减法和除法,但比之乘法来說,它們所费的精力时間可以认为是微不足道的。)例如在$n=10$的情形大約需要作3.6×10^8次乘法。如果用台式計算机,假定每小时能作100次乘法,按每天5小时,每年200天,即按每年工作1000小时計算,一个人一年可以作10^5次乘法。按照这样的速度作3.6×10^8次乘法就需要3600年的时間。显然,这是完全沒有实际意义的。

基于上述的理由,作为联立1次方程的解法,必需考虑另外的方法。过去所用的数値解法大体上可分为消去法和迭代法两类。本章在§1,§2中討論消去法,而在§3中討論迭代法。在§4中我們将叙述一个兼有上述两种方法特征的新的方法。

消去法是由給定的方程組出发,有时将方程乘以常数,或者将两个以上的方程相加相减,逐步消去未知数,最后导出只含一个未知数的方程。这些中間計算相当于把以系数(和"右边")为分量的向量乘以常数,或者将几个向量相加减的运算。因此可以稍为抽象地叙述如下:

用常数 c 乘向量 $(a_1, a_2, \cdots, a_n)$ 而作向量 $(ca_1, ca_2, \cdots, ca_n)$，或由向量 $(a_1, a_2, \cdots, a_n)$ 减去向量 $(b_1, b_2, \cdots, b_n)$ 的 k 倍而作向量 $(a_1-kb_1, a_2-kb_2, \cdots, a_n-kb_n)$ 的运算都是綫性(linear)运算。

例如，以矩陣

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{pmatrix} \tag{1.1}$$

的各行为单位，进行上述的綫性运算，就可以把它化成如下的形状：

$$\begin{pmatrix} 1 & a'_{12} & a'_{13} & a'_{14} & a'_{15} \\ 0 & a'_{22} & a'_{23} & a'_{24} & a'_{25} \\ 0 & a'_{32} & a'_{33} & a'_{34} & a'_{35} \end{pmatrix}. \tag{1.2}$$

为此，首先用 a_{11} 除(1.1)的第1行(或者也就是用 $1/a_{11}$ 乘)得到

$$a'_{1j}=a_{1j}/a_{11} \quad (j=2, \cdots, 5), \tag{1.3}$$

其次，对于其他各行計算

$$a'_{ij}=a_{ij}-\frac{a_{i1}a_{1j}}{a_{11}}=a_{ij}-a_{i1}a'_{1j} \quad (i=2, 3;\ j=2, \cdots, 5) \tag{1.4}$$

就可以了。这样的計算叫做以元素 a_{11} 为主元素(pivot)的扫除法(sweep-out)[1]。

一般来說，对于 m 行 n 列 $(m<n)$ 的矩陣 (a_{ij})，以 $a_{i^*j^*}$ 为主元素的扫除法就是首先关于第 i^* 行計算

$$a'_{i^*j}=a_{i^*j}/a_{i^*j^*} \quad (j=1, \cdots, n), \tag{1.5}$$

然后关于其他各行計算

$$a'_{ij}=a_{ij}-\frac{a_{ij^*}a_{i^*j}}{a_{i^*j^*}}=a_{ij}-a_{ij^*}a'_{i^*j} \text{[2]} \quad (i\neq i^*, j=1, \cdots, n). \tag{1.6}$$

---

[1] 这方法在我国的文献中通常称为 Gauss 方法。——譯者注

[2] 关于用穿孔卡計算机进行这种計算的問題，請参看本丛书中森口著《穿孔卡計算机》§ 11.

这样就得到 $a'_{i^*j^*}=1$, $a'_{ij^*}=0(i\neq i^*)$，因此所得矩陣 (a'_{ij}) 的第 j^* 列成为单位向量。

任何元素只要不等于零都可以取作主元素，按照不同的情况可以有种种的取法。在联立一次方程的解法中，通常由左上角开始逐漸向右下角移动，最后到达第 m 行第 m 列为止。这样，右面就出現了 m 阶的单位矩陣。(在以 $a_{i^*j^*}$ 为主元素的扫除法中，第 i^* 行的元素为 0 的列是不变的，因此，在以上的計算中，一旦成为单位向量的列，以后就不再变化，直到最后仍旧保持为单位向量。)

在綫性规划(参看本丛书森口、宮下著《綫性規划》)的单純形計算法中，把对应于取作基底的变量的列作为第 j^* 列，对应于要消去的变量的行作为第 i^* 行，而以 $a_{i^*j^*}$ 为主元素。并且由(1.5)計算下一个单純形表的"新出現的行"，而由(1.6)計算其他的各行。

例 1 解联立方程：

$$\left.\begin{aligned}2x+3y+4z&=6,\\3x+5y+2z&=5,\\4x+3y+30z&=32.\end{aligned}\right\}\tag{1.7}$$

按照表 1.1 計算，得到解 $x=-13$, $y=8$, $z=2$.

[注 1] 表 1.1 最上一行的 x, y, z = 可以看作是由(1.7)的三个方程提出来的共同因素。右边的"1"可以看作是和表中的 6 或 5 相乘的对象。行(1)表示 2 乘以 x, 3 乘以 y, 4 乘以 z 而后相加等于 6 乘以 1. 对于以下的各行，也可以作同样的解釋。特别是，行(10)表示 $1\cdot x+0\cdot y+0\cdot z=-13\cdot 1$，即 $x=-13$，类似地，行(11)表示 $y=8$，行(12)表示 $z=2$.

[注 2] 通常在初等数学中讲授的消去法，如果用表的形式表示，则以表 1.2 較为恰当。和这种方法比較，表 1.1 具有沒有"逆行部分"的特征。消去法終了时一下就得到了解的全部。从乘法的次数来看，两者大致相同，約有 n^3 次左右，但計算的方案可說表 1.1 比較簡单，因为它是步驟相同的程序的重复。基于这样的理由，在机械計算中有更多地使用表 1.1 中的方法的傾向。

一般說来，如果向量 $(a_1, \cdots, a_n)$ 滿足一次关系式

表1.1　联立方程解法(扫除法)

行	x	y	z	= 1	説　明
(1)	2	3	4	6	表示(1.7).
(2)	3	5	2	5	
(3)	4	3	30	32	
(4)	1	1.5	2	3	(1)/2
(5)	0	0.5	−4	−4	(2)−(4)×3
(6)	0	−3	22	20	(3)−(4)×4
(7)	1	0	14	15	(4)−(8)×1.5
(8)	0	1	−8	−8	(5)/0.5
(9)	0	0	−2	−4	(6)+(8)×3
(10)	1	0	0	−13	(7)−(12)×14
(11)	0	1	0	8	(8)+(12)×8
(12)	0	0	1	2	(9)/(−2)

表1.2　通常的消去法

行	x	y	z	= 1	説　明	
(1)	2	3	4	6	表示(1.7)	
(2)	3	5	2	5		
(3)	4	3	30	32		
(4)	1	1.5	2	3	(1)/2	前进部分
(5)	0	0.5	−4	−4	(2)−(4)×3	
(6)	0	−3	22	20	(3)−(4)×4	
(7)	0	1	−8	−8	(5)/0.5	
(8)	0	0	−2	−4	(6)+(8)×3	
(9)	0	0	1	2	(9)/(−2)	
(10)	0	1	0	8	(7)+(9)×8	逆行部分
(11)	1	1.5	0	−1	(4)−(9)×2	
(12)	1	0	0	−13	(11)−(10)×1.5	

$$a_1\lambda_1+\cdots+a_n\lambda_n=0, \tag{1.8}$$

那么,用 c 乘后所得的向量 $(ca_1, \cdots, ca_r)$ 也滿足同样的关系式

$$(ca_1)\lambda_1+\cdots+(ca_n)\lambda_n=0. \tag{1.9}$$

如果还有另一向量 $(b_1, \cdots, b_n)$ 也滿足同一关系式

$$b_1\lambda_1+\cdots+b_n\lambda_n=0, \tag{1.10}$$

那么, $(a_1-kb_1, \cdots, a_n-kb_n)$ 也滿足同一关系式

$$(a_1-kb_1)\lambda_1+\cdots+(a_n-kb_n)\lambda_n=0. \tag{1.11}$$

因此,形如(1.8)的关系式在綫性运算的过程中經常保持不变。由此导出如下的定理:

定理 由矩陣 (a_{ij}) 反复使用扫除法得到矩陣 (b_{ij}) 时,如果原来的矩陣的列之間有如下的一次关系成立:

$$\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}\lambda_1+\cdots+\begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}\lambda_n=0, \tag{1.12}$$

那么,关于矩陣 (b_{ij}) 也有同样的关系式成立:

$$\begin{pmatrix} b_{11} \\ \vdots \\ b_{m1} \end{pmatrix}\lambda_1+\cdots+\begin{pmatrix} b_{1n} \\ \vdots \\ b_{mn} \end{pmatrix}\lambda_n=0. \tag{1.13}$$

例 2 对于表 1.1 开首的矩陣(行(1),(2),(3)的部分),下列关系式成立:

$$\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}x+\begin{pmatrix} 3 \\ 5 \\ 3 \end{pmatrix}y+\begin{pmatrix} 4 \\ 2 \\ 30 \end{pmatrix}z=\begin{pmatrix} 6 \\ 5 \\ 32 \end{pmatrix}.$$

对于最后的矩陣,也应該有同样的关系式成立:

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}x+\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}y+\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}z=\begin{pmatrix} -13 \\ 8 \\ 2 \end{pmatrix}.$$

換句話說,如果 x, y, z 是(1.7)的解,那就必需是 $x=-13$, $y=8$, $z=2$。这样,在最后的矩陣中得到的就是解。

例 3 在表 1.1 开首的矩陣的右边加上单位矩陣,再在它的右边加上"总

計”列，然后如对表 1.1 那样进行(扫除法)計算就得到如下的表 1.3.

表 1.3　联立一次方程的解及其逆矩陣的計算

行	x	y	z	1	C_1	C_2	C_3	总計	說　　明
(1)	2	3	4	6	1	0	0	16	(1.7)，单位矩陣，总計
(2)	3	5	2	5	0	1	0	16	
(3)	4	3	30	32	0	0	1	70	
(4)	1	1.5	2	3	0.5	0	0	8	(1)/2
(5)	0	0.5	−4	−4	−1.5	1	0	−8	(2)−(4)×3
(6)	0	−3	22	20	−2	0	1	28	(3)−(4)×4
(7)	1	0	14	15	5	−3	0	32	(4)−(8)×1.5
(8)	0	1	−8	−8	−3	2	0	−16	(5)/0.5
(9)	0	0	−2	−4	−11	6	1	−10	(6)+(8)×3
(10)	1	0	0	−13	−72	39	7	−38	(7)−(12)×14
(11)	0	1	0	8	41	−22	−4	24	(8)+(12)×8
(12)	0	0	1	2	5.5	−3	−0.5	5	(9)/(−2)

在这种情形，如果在开首的矩陣中

$$\begin{pmatrix}2\\3\\4\end{pmatrix}x+\begin{pmatrix}3\\5\\3\end{pmatrix}y+\begin{pmatrix}4\\2\\30\end{pmatrix}z=\begin{pmatrix}1\\0\\0\end{pmatrix}u+\begin{pmatrix}0\\1\\0\end{pmatrix}v+\begin{pmatrix}0\\0\\1\end{pmatrix}w$$

成立，那么在最后的矩陣中

$$\begin{pmatrix}1\\0\\0\end{pmatrix}x+\begin{pmatrix}0\\1\\0\end{pmatrix}y+\begin{pmatrix}0\\0\\1\end{pmatrix}z=\begin{pmatrix}-72\\41\\5.5\end{pmatrix}u+\begin{pmatrix}39\\-22\\-3\end{pmatrix}v+\begin{pmatrix}7\\-4\\-0.5\end{pmatrix}w$$

也必需成立。換句話說，如果

$$\begin{pmatrix}2&3&4\\3&5&2\\4&3&30\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}u\\v\\w\end{pmatrix},$$

那么

$$\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}-72&39&7\\41&-22&-4\\5.5&-3&-0.5\end{pmatrix}\begin{pmatrix}u\\v\\w\end{pmatrix}.$$

而这就表示

$$\begin{pmatrix} -72 & 39 & 7 \\ 41 & -22 & -4 \\ 5.5 & -3 & -0.5 \end{pmatrix}$$

$$=\begin{pmatrix} 2 & 3 & 4 \\ 3 & 5 & 2 \\ 4 & 3 & 30 \end{pmatrix}^{-1}.$$

这样，在最后矩陣的 C_1, C_2, C_3 部分就得到原来"左边"的矩陣的逆矩陣。

此外，对于总計列，在开首的矩陣中下列关系式成立：

$$\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}+\begin{pmatrix} 3 \\ 5 \\ 3 \end{pmatrix}+\begin{pmatrix} 4 \\ 2 \\ 30 \end{pmatrix}+\begin{pmatrix} 6 \\ 5 \\ 32 \end{pmatrix}+\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}+\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}+\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$=\begin{pmatrix} 16 \\ 16 \\ 70 \end{pmatrix},$$

因此，在以下的每一行中，左面的元素之和必需等于总計列的对应元素。例如，在行(10)中

$$1+0+0-13-72+39+7=-38$$

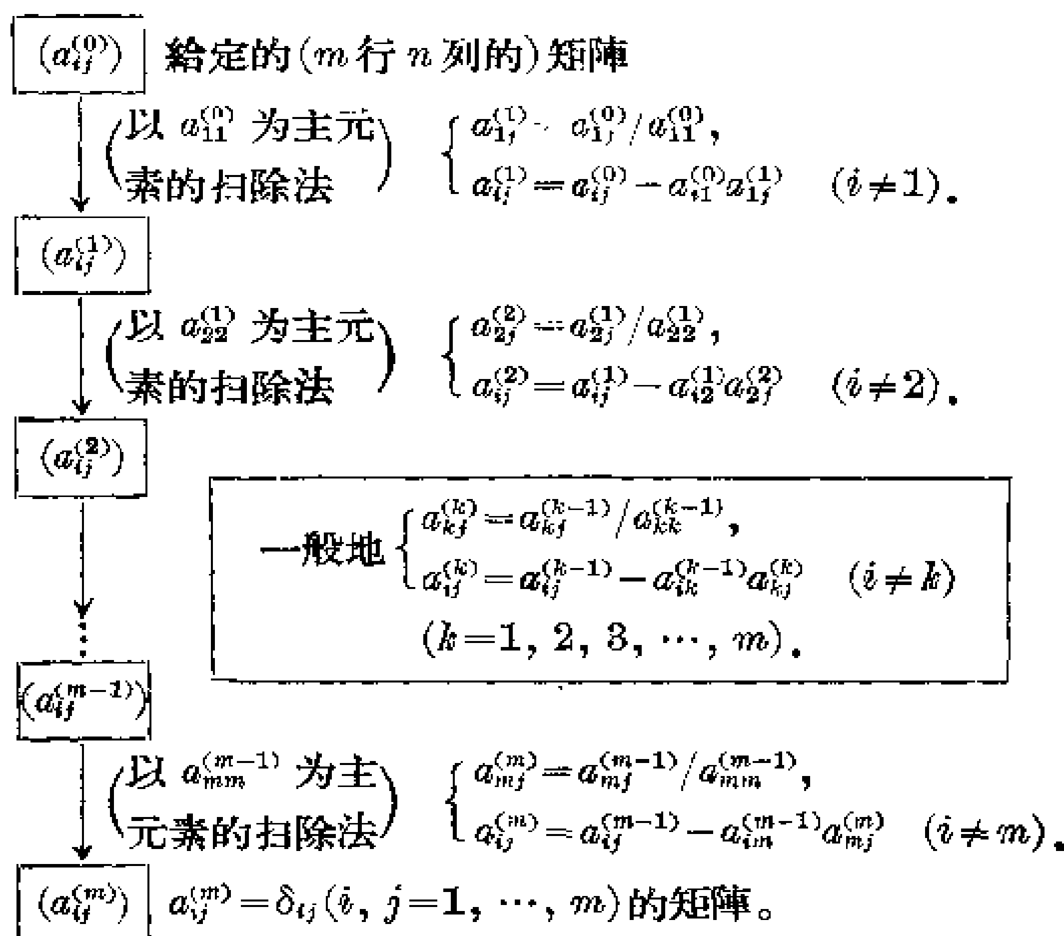
应当成立。这一事实有助于在計算过程中作驗算。

［注1］一般来說，如果在开首的矩陣中有 (A, B, I) 的部分，(A 是方陣，B 是任意矩陣，I 是单位矩陣)，在反复实行扫除法計算的过程中，变成了 (I, D, C) 的形状，那么，C 是 A 的逆矩陣而 D 等于 $A^{-1}B$.

［注2］在作綫性运算时希望一定要如上面那样加入"和的驗算(sum check)"，因为在发生了錯誤时，能及早发現糾正是很重要的。

［注3］上述的定理在綫性規划中单純形表的利用上也可以便利地应用(参看森口著：綫性規划入門，日本科技联，1957，§10)。

以上說明的扫除法的計算程序可以归納成如下的流动图(flowchart)：



其中假定了每一个主元素 $a_{kk}^{(k-1)}(k=1, 2, \cdots, m)$都不等于0.

此外，系数矩陣$(a_{ij}^{(0)};\ i, j=1, \cdots, m)$的行列式的値，可以表示成在計算过程中使用的一切主元素的乘积，即

$$\begin{vmatrix} a_{11}^{(0)} & \cdots & a_{1m}^{(0)} \\ \vdots & & \vdots \\ a_{m1}^{(0)} & \cdots & a_{mm}^{(0)} \end{vmatrix} = a_{11}^{(0)}a_{22}^{(1)}\cdots a_{mm}^{(m-1)}. \tag{1.14}$$

証明 对左边的行列式应用上述的扫除法，一般地在第k次的扫除法中，用$a_{kk}^{(k-1)}$除第k行时，行列式的値成为$1/a_{kk}^{(k-1)}$倍，而在改变其他各行的計算中行列式的値保持不变。而且最后得到的是单位矩陣的行列式，因此它的値等于1. 逆推回去就得到(1.14)。 証毕

§2 記录的簡化(紧凑法，簡化 Doolittle 法，平方根法)

前节所叙述的扫除法中，在m行n列的矩陣的情形，需要記

录的数值总共有 $mn(m+1)$ 个($m+1$ 个 m 行 n 列的矩陣)。但是,用台式計算机进行計算时,并不需要記录这些数值的全部。例如,在 $m=4$, $n=9$ 的情形,只要有表 2.1 中那样的記录就可以进行計算。(表中記号的意义与前节末的表相同。)这就是所謂**紧凑法**(compact method)或 Crout 法。

表 2.1　紧凑法(Crout 法)

行＼列	1	2	3	4	5	6	7	8	9	总　計
(1)	$a_{11}^{(0)}$	$a_{12}^{(0)}$	$a_{13}^{(0)}$	$a_{14}^{(0)}$	$a_{15}^{(0)}$	$a_{16}^{(0)}$	$a_{17}^{(0)}$	$a_{18}^{(0)}$	$a_{19}^{(0)}$	$a_{1T}^{(0)}$
(2)	$a_{21}^{(0)}$	$a_{22}^{(0)}$	$a_{23}^{(0)}$	$a_{24}^{(0)}$	$a_{25}^{(0)}$	$a_{26}^{(0)}$	$a_{27}^{(0)}$	$a_{28}^{(0)}$	$a_{29}^{(0)}$	$a_{2T}^{(0)}$
(3)	$a_{31}^{(0)}$	$a_{32}^{(0)}$	$a_{33}^{(0)}$	$a_{34}^{(0)}$	$a_{35}^{(0)}$	$a_{36}^{(0)}$	$a_{37}^{(0)}$	$a_{38}^{(0)}$	$a_{39}^{(0)}$	$a_{3T}^{(0)}$
(4)	$a_{41}^{(0)}$	$a_{42}^{(0)}$	$a_{43}^{(0)}$	$a_{44}^{(0)}$	$a_{45}^{(0)}$	$a_{46}^{(0)}$	$a_{47}^{(0)}$	$a_{48}^{(0)}$	$a_{49}^{(0)}$	$a_{4T}^{(0)}$
(5)	$a_{11}^{(0)}$	$a_{12}^{(1)}$	$a_{13}^{(1)}$	$a_{14}^{(1)}$	$a_{15}^{(1)}$	$a_{16}^{(1)}$	$a_{17}^{(1)}$	$a_{18}^{(1)}$	$a_{19}^{(1)}$	$a_{1T}^{(1)}$
(6)	$a_{21}^{(0)}$	$a_{22}^{(1)}$	$a_{23}^{(2)}$	$a_{24}^{(2)}$	$a_{25}^{(2)}$	$a_{26}^{(2)}$	$a_{27}^{(2)}$	$a_{28}^{(2)}$	$a_{29}^{(2)}$	$a_{2T}^{(2)}$
(7)	$a_{31}^{(0)}$	$a_{32}^{(1)}$	$a_{33}^{(2)}$	$a_{34}^{(3)}$	$a_{35}^{(3)}$	$a_{36}^{(3)}$	$a_{37}^{(3)}$	$a_{38}^{(3)}$	$a_{39}^{(3)}$	$a_{3T}^{(3)}$
(8)	$a_{41}^{(0)}$	$a_{42}^{(1)}$	$a_{43}^{(2)}$	$a_{44}^{(3)}$	$a_{45}^{(4)}$	$a_{46}^{(4)}$	$a_{47}^{(4)}$	$a_{48}^{(4)}$	$a_{49}^{(4)}$	$a_{4T}^{(4)}$
(9)	1	0	0	0	$a_{15}^{(4)}$	$a_{16}^{(4)}$	$a_{17}^{(4)}$	$a_{18}^{(4)}$	$a_{19}^{(4)}$	$a_{1T}^{(4)}$
(10)	0	1	0	0	$a_{25}^{(4)}$	$a_{26}^{(4)}$	$a_{27}^{(4)}$	$a_{28}^{(4)}$	$a_{29}^{(4)}$	$a_{2T}^{(4)}$
(11)	0	0	1	0	$a_{35}^{(4)}$	$a_{36}^{(4)}$	$a_{37}^{(4)}$	$a_{38}^{(4)}$	$a_{39}^{(4)}$	$a_{3T}^{(4)}$
(12)	0	0	0	1	$a_{45}^{(4)}$	$a_{46}^{(4)}$	$a_{47}^{(4)}$	$a_{48}^{(4)}$	$a_{49}^{(4)}$	$a_{4T}^{(4)}$

表 2.1 的說明　(1)～(4)行是給定的矩陣$(a_{ij}^{(0)})$. 总計列是令

$$a_{iT}^{(0)}=\sum_{j=1}^{9}a_{ij}^{(0)}$$

而計算的。

(5)～(8)行是"前进部分"。首先,$a_{11}^{(0)}$, $a_{21}^{(0)}$, $a_{31}^{(0)}$, $a_{41}^{(0)}$ 是由上面完全不变地移下来的。其次,$a_{12}^{(1)}$, $a_{13}^{(1)}$, $\cdots$, $a_{19}^{(1)}$, $a_{1T}^{(1)}$ 是由公式

$$a_{1j}^{(1)}=a_{1j}^{(0)}/a_{11}^{(0)}\quad(j=2,\ 3,\ \cdots,\ 9,\ T)$$

計算的。再一步是由公式

$$a_{i2}^{(1)}=a_{i2}^{(0)}-a_{i1}^{(0)}a_{12}^{(1)}\quad(i=2,\ 3,\ 4)$$

計算 $a_{22}^{(1)}$, $a_{32}^{(1)}$, $a_{42}^{(1)}$ (这里需要对 $a_{12}^{(1)}$ 已經計算出来了)。在这以后就是按照

下列公式逐步进行計算：

$$a_{2j}^{(2)}=\{a_{2j}^{(0)}-a_{21}^{(0)}a_{1j}^{(1)}\}/a_{22}^{(1)}\quad (j=3,\ 4,\ \cdots,\ 9,\ T),$$
$$a_{i3}^{(2)}=a_{i3}^{(0)}-a_{i1}^{(0)}a_{13}^{(1)}-a_{i2}^{(1)}a_{23}^{(2)}\quad (i=3,\ 4),$$
$$a_{3j}^{(3)}=\{a_{3j}^{(0)}-a_{31}^{(0)}a_{1j}^{(1)}-a_{32}^{(1)}a_{2j}^{(2)}\}/a_{33}^{(2)}\quad (j=4,\ \cdots,\ 9,\ T),$$
$$a_{44}^{(3)}=a_{44}^{(0)}-a_{41}^{(0)}a_{14}^{(1)}-a_{42}^{(1)}a_{24}^{(2)}-a_{43}^{(2)}a_{34}^{(3)},$$
$$a_{4j}^{(4)}=\{a_{4j}^{(0)}-a_{41}^{(0)}a_{1j}^{(1)}-a_{42}^{(1)}a_{2j}^{(2)}-a_{43}^{(2)}a_{3j}^{(3)}\}/a_{44}^{(3)}\quad (j=5,\ \cdots,\ 9,\ T).$$

驗算根据

$$a_{iT}^{(i)}=1+a_{i,i+1}^{(i)}+a_{i,i+2}^{(i)}+\cdots+a_{i9}^{(i)}\quad (i=1,\ 2,\ 3,\ 4)$$

进行。

(9)～(12)行是"逆行部分"。1～4列的部分填写单位矩陣。(12)行是将(8)行不变地移下来的。(11)行是由(7)行和(12)行按公式

$$a_{3j}^{(4)}=a_{3j}^{(3)}-a_{34}^{(3)}a_{4j}^{(4)}\quad (j=5,\ \cdots,\ 9,\ T)$$

求得的。(10)行是由(6)行和(11)，(12)行按公式

$$a_{2j}^{(4)}=a_{2j}^{(2)}-a_{23}^{(2)}a_{3j}^{(4)}-a_{24}^{(2)}a_{4j}^{(4)}\quad (j=5,\ \cdots,\ 9,\ T)$$

求得的，而(9)行是由(5)行和(10)，(11)，(12)行按公式

$$a_{1j}^{(4)}=a_{1j}^{(1)}-a_{12}^{(1)}a_{2j}^{(4)}-a_{13}^{(1)}a_{3j}^{(4)}-a_{14}^{(1)}a_{4j}^{(4)}\quad (j=5,\ \cdots,\ 9,\ T)$$

求得的。(5)，(6)，(7)，(8) 行的阶梯綫右面的部分实际上是表 2.2 的一部分，因此可以将它的 (8)′ 行移作 (12) 行，由(7)′ 减去(12)的 $a_{34}^{(3)}$ 倍而得到(11)，由(6)′ 减去(11)的 $a_{23}^{(2)}$ 倍，和(12)的 $a_{24}^{(2)}$ 倍得到(10)，由(5)′ 减去(10)的 $a_{12}^{(1)}$ 倍，(11)的 $a_{13}^{(1)}$ 倍，(12)的 $a_{14}^{(1)}$ 倍得到(9)。

表 2.2 (5)～(8)行的說明

行＼列	1	2	3	4	5	6	7	8	9	总 計
(5)′	1	$a_{12}^{(1)}$	$a_{13}^{(1)}$	$a_{14}^{(1)}$	$a_{15}^{(1)}$	$a_{16}^{(1)}$	$a_{17}^{(1)}$	$a_{18}^{(1)}$	$a_{19}^{(1)}$	$a_{1T}^{(1)}$
(6)′	0	1	$a_{23}^{(2)}$	$a_{24}^{(2)}$	$a_{25}^{(2)}$	$a_{26}^{(2)}$	$a_{27}^{(2)}$	$a_{28}^{(2)}$	$a_{29}^{(2)}$	$a_{2T}^{(2)}$
(7)′	0	0	1	$a_{34}^{(3)}$	$a_{35}^{(3)}$	$a_{36}^{(3)}$	$a_{37}^{(3)}$	$a_{38}^{(3)}$	$a_{39}^{(3)}$	$a_{3T}^{(3)}$
(8)′	0	0	0	1	$a_{45}^{(4)}$	$a_{46}^{(4)}$	$a_{47}^{(4)}$	$a_{48}^{(4)}$	$a_{49}^{(4)}$	$a_{4T}^{(4)}$

[注 1] 这一方法的优点在于，用台式計算机作形如 $A-BC-DE$ 和

$(A-BC-DE)/F$ 的計算时,可以不必在紙上記录計算过程中的数值而用連續的操作来完成它。这种中間記录的省略,不只可以节約紙張、墨水和劳动力,也减少了将数值由計算机移到紙上,再由紙上移到計算机上的次数,因而相应地减少了发生錯誤的机会(一般說来,錯誤多半是由人引起的)。

[注 2] 在作表 2.1 的計算时,不必一一按照"說明"部分所列举的公式进行,只要由表上取用必要的数值,按"規則"进行就可以了。把这規則写成程序的形状是一个很好的习題。(为了計算前进部分的某一数値所必要的数值只不过是,最初的矩陣中和它对应的元素,前进部分自身中含有所求数值位置的行与列中已經写出的元素而已。)

例 1 用紧凑法解表 1.3 的問題,就得到形如表 2.3 的記录。

表 2.3 紧凑法的例子

行＼列	1	2	3	4	5	6	7	总 計
(1)	2	3	4	6	1	0	0	16
(2)	3	5	2	5	0	1	0	16
(3)	4	3	30	32	0	0	1	70
(4)	2	1.5	2	3	0.5	0	0	8
(5)	3	0.5	−8	−8	−3	2	0	−16
(6)	4	−3	−2	2	5.5	−3	−0.5	5
(7)	1	0	0	−13	−72	39	7	−38
(8)	0	1	0	8	41	−22	−4	24
(9)	0	0	1	2	5.5	−3	−0.5	5

特别是,如果系数矩陣是对称的,那么,由此用扫除法导出的矩陣也反映着这种对称性[一般地,$a_{ij}^{(k)}=a_{ji}^{(k)}\ (k\leqslant i<j\leqslant m)$],利用这一点可以依据如表 2.4 那样的記录进行計算。这一方法叫做**簡化 Doolittle 法**,在統計分析的正規方程解法等問題中是經常使用的。

表 2.4　簡化 Doolittle 法

行＼列	1	2	3	4	5	6	7	8	9	總計	說　明	
(1)	$a_{11}^{(0)}$	$a_{12}^{(0)}$	$a_{13}^{(0)}$	$a_{14}^{(0)}$	$a_{15}^{(0)}$	$a_{16}^{(0)}$	$a_{17}^{(0)}$	$a_{18}^{(0)}$	$a_{19}^{(0)}$	$a_{17}^{(0)}$	給定的矩陣。有＊号的地方由于对称性可以不必写出，但是，例如要注意：$a_{37}^{(0)}=a_{13}^{(0)}+a_{23}^{(0)}+a_{33}^{(0)}+a_{34}^{(0)}+\cdots+a_{39}^{(0)}$.	
(2)	*	$a_{22}^{(0)}$	$a_{23}^{(0)}$	$a_{24}^{(0)}$	$a_{25}^{(0)}$	$a_{26}^{(0)}$	$a_{27}^{(0)}$	$a_{28}^{(0)}$	$a_{29}^{(0)}$	$a_{27}^{(0)}$		
(3)	*	*	$a_{33}^{(0)}$	$a_{34}^{(0)}$	$a_{35}^{(0)}$	$a_{36}^{(0)}$	$a_{37}^{(0)}$	$a_{38}^{(0)}$	$a_{39}^{(0)}$	$a_{37}^{(0)}$		
(4)	*	*	*	$a_{44}^{(0)}$	$a_{45}^{(0)}$	$a_{46}^{(0)}$	$a_{47}^{(0)}$	$a_{48}^{(0)}$	$a_{49}^{(0)}$	$a_{47}^{(0)}$		
(5)	$a_{11}^{(0)}$	$a_{12}^{(0)}$	$a_{13}^{(0)}$	$a_{14}^{(0)}$	$a_{15}^{(0)}$	$a_{16}^{(0)}$	$a_{17}^{(0)}$	$a_{18}^{(0)}$	$a_{19}^{(0)}$	$a_{17}^{(0)}$	行(1)的复写	前进部分
(5)′	1	$a_{12}^{(1)}$	$a_{13}^{(1)}$	$a_{14}^{(1)}$	$a_{15}^{(1)}$	$a_{16}^{(1)}$	$a_{17}^{(1)}$	$a_{18}^{(1)}$	$a_{19}^{(1)}$	$a_{17}^{(1)}$	$(5)/a_{11}^{(0)}$	
(6)		$a_{22}^{(1)}$	$a_{23}^{(1)}$	$a_{24}^{(1)}$	$a_{25}^{(1)}$	$a_{26}^{(1)}$	$a_{27}^{(1)}$	$a_{28}^{(1)}$	$a_{29}^{(1)}$	$a_{27}^{(1)}$	$(2)-(5)\times a_{12}^{(1)}$	
(6)′		1	$a_{23}^{(2)}$	$a_{24}^{(2)}$	$a_{25}^{(2)}$	$a_{26}^{(2)}$	$a_{27}^{(2)}$	$a_{28}^{(2)}$	$a_{29}^{(2)}$	$a_{27}^{(2)}$	$(6)/a_{22}^{(1)}$	
(7)			$a_{33}^{(2)}$	$a_{34}^{(2)}$	$a_{35}^{(2)}$	$a_{36}^{(2)}$	$a_{37}^{(2)}$	$a_{38}^{(2)}$	$a_{39}^{(2)}$	$a_{37}^{(2)}$	$(3)-(5)\times a_{13}^{(1)}$ $-(6)\times a_{23}^{(2)}$	
(7)′			1	$a_{34}^{(3)}$	$a_{35}^{(3)}$	$a_{36}^{(3)}$	$a_{37}^{(3)}$	$a_{38}^{(3)}$	$a_{39}^{(3)}$	$a_{37}^{(3)}$	$(7)/a_{33}^{(2)}$	
(8)				$a_{44}^{(3)}$	$a_{45}^{(3)}$	$a_{46}^{(3)}$	$a_{47}^{(3)}$	$a_{48}^{(3)}$	$a_{49}^{(3)}$	$a_{47}^{(3)}$	$(4)-(5)\times a_{14}^{(1)}$ $-(6)\times a_{24}^{(2)}$ $-(7)\times a_{34}^{(3)}$	
(8)′				1	$a_{45}^{(4)}$	$a_{46}^{(4)}$	$a_{47}^{(4)}$	$a_{48}^{(4)}$	$a_{49}^{(4)}$	$a_{47}^{(4)}$	$(8)/a_{44}^{(3)}$	
(9)	1				$a_{15}^{(4)}$	$a_{16}^{(4)}$	$a_{17}^{(4)}$	$a_{18}^{(4)}$	$a_{19}^{(4)}$	$a_{17}^{(4)}$	$(5)'-(12)\times a_{14}^{(1)}$ $-(11)\times a_{13}^{(1)}$ $-(10)\times a_{12}^{(1)}$	逆行部分
(10)		1			$a_{25}^{(4)}$	$a_{26}^{(4)}$	$a_{27}^{(4)}$	$a_{28}^{(4)}$	$a_{29}^{(4)}$	$a_{27}^{(4)}$	$(6)'-(12)\times a_{24}^{(2)}$ $-(11)\times a_{23}^{(2)}$	
(11)			1		$a_{35}^{(4)}$	$a_{36}^{(4)}$	$a_{37}^{(4)}$	$a_{38}^{(4)}$	$a_{39}^{(4)}$	$a_{37}^{(4)}$	$(7')-(12)\times a_{34}^{(3)}$	
(12)				1	$a_{45}^{(4)}$	$a_{46}^{(4)}$	$a_{47}^{(4)}$	$a_{48}^{(4)}$	$a_{49}^{(4)}$	$a_{47}^{(4)}$	(8)′的复写	

［注］　在这种情形也同样，如果把計算的规則用表中数值的位置陈述出来是便利的(位置原理1)。

在表 2.4 的前进部分中，(5)，(5)′；(6)，(6)′；(7)，(7)′；(8)，(8)′两行两行地成对，在每一对中下面的行是用常数(最靠左面的不等于 0 的元素 $a_{kk}^{(k-1)}$)除上面的行而得到的。代替这样的行对，如果使用对应元素的几何平均数，卽将上面的行用

$$s_k=\sqrt{a_{kk}^{(k-1)}}$$

除得的結果，就可以按照表 2.5 所記录的那样进行計算。这方法叫做**平方根法**，在用台式計算机解元数很多的正規方程时是常常使用的。

表2.5 平 方 根 法

行＼列	1	2	3	4	5	6	7	8	9	总計	說明	
(1)	$a_{11}^{(0)}$	$a_{12}^{(0)}$	$a_{13}^{(0)}$	$a_{14}^{(0)}$	$a_{15}^{(0)}$	$a_{16}^{(0)}$	$a_{17}^{(0)}$	$a_{18}^{(0)}$	$a_{19}^{(0)}$	$a_{1T}^{(0)}$	給定的矩陣。有 * 号的地方根据对称性略去。	
(2)	*	$a_{22}^{(0)}$	$a_{23}^{(0)}$	$a_{24}^{(0)}$	$a_{25}^{(0)}$	$a_{26}^{(0)}$	$a_{27}^{(0)}$	$a_{28}^{(0)}$	$a_{29}^{(0)}$	$a_{2T}^{(0)}$		
(3)	*	*	$a_{33}^{(0)}$	$a_{34}^{(0)}$	$a_{35}^{(0)}$	$a_{36}^{(0)}$	$a_{37}^{(0)}$	$a_{38}^{(0)}$	$a_{39}^{(0)}$	$a_{3T}^{(0)}$	$a_{iT}^{(0)}=a_{1i}^{(0)}+\cdots+a_{i-1,i}^{(0)}+a_{ii}^{(0)}$	
(4)	*	*	*	$a_{44}^{(0)}$	$a_{45}^{(0)}$	$a_{46}^{(0)}$	$a_{47}^{(0)}$	$a_{48}^{(0)}$	$a_{49}^{(0)}$	$a_{4T}^{(0)}$	$+a_{i,i+1}^{(0)}+\cdots+a_{i9}^{(0)}$	
(5)	s_1	g_{12}	g_{13}	g_{14}	g_{15}	g_{16}	g_{17}	g_{18}	g_{19}	g_{1T}	$s_1=\sqrt{a_{11}^{(0)}}$，其余的$(1)/s_1$	前进部分
(6)		s_2	g_{23}	g_{24}	g_{25}	g_{26}	g_{27}	g_{28}	g_{29}	g_{2T}	$s_2=\sqrt{a_{22}^{(0)}-g_{12}^2}$，其余的$\{(2)-(5)\times g_{12}\}/s_2$	
(7)			s_3	g_{34}	g_{35}	g_{36}	g_{37}	g_{38}	g_{39}	g_{3T}	$s_3=\sqrt{a_{33}^{(0)}-g_{13}^2-g_{23}^2}$，其余的$\{(3)-(5)\times g_{13}-(6)\times g_{23}\}/s_3$	
(8)				s_4	g_{45}	g_{46}	g_{47}	g_{48}	g_{49}	g_{4T}	$s_4=\sqrt{a_{44}^{(0)}-g_{14}^2-g_{24}^2-g_{34}^2}$ 其余的$\{(4)-(5)\times g_{14}-(6)\times g_{24}-(7)\times g_{34}\}/s_4$	
(9)	1				$a_{15}^{(4)}$	$a_{16}^{(4)}$	$a_{17}^{(4)}$	$a_{18}^{(4)}$	$a_{19}^{(4)}$	$a_{1T}^{(4)}$	$\{(5)-(12)\times g_{14}-(11)\times g_{13}-(10)\times g_{12}\}/s_1$	逆行部分
(10)		1			$a_{25}^{(4)}$	$a_{26}^{(4)}$	$a_{27}^{(4)}$	$a_{28}^{(4)}$	$a_{29}^{(4)}$	$a_{2T}^{(4)}$	$\{(6)-(12)\times g_{24}-(11)\times g_{23}\}/s_2$	
(11)			1		$a_{35}^{(4)}$	$a_{36}^{(4)}$	$a_{37}^{(4)}$	$a_{38}^{(4)}$	$a_{39}^{(4)}$	$a_{3T}^{(4)}$	$\{(7)-(12)\times g_{34}\}/s_3$	
(12)				1	$a_{45}^{(4)}$	$a_{46}^{(4)}$	$a_{47}^{(4)}$	$a_{48}^{(4)}$	$a_{49}^{(4)}$	$a_{4T}^{(4)}$	$(8)/s_4$	

[注] 平方根法可以看作是几何学中把給定向量系"标准正交化"的手續(参看本丛书森口著《統計分析》p.44 的例子)。

关于綫性运算，特别是本节所討論的問題詳見 Dwyer 的书(文献[1])。

§3　迭代法(Gauss-Seidel 方法)

如果給定的方程組 $\sum a_{ij}x_j=b_i$ 的系数矩陣接近于对角矩陣(卽其他元素比之对角綫上的元素較小)，将方程組改写成

$$\left.\begin{aligned} x_1&=(b_1-a_{12}x_2-\cdots-a_{1n}x_n)/a_{11},\\ &\cdots\cdots\cdots\cdots\cdots\cdots\\ x_n&=(b_n-a_{n1}x_1-\cdots-a_{n,n-1}x_{n-1})/a_{nn} \end{aligned}\right\}\qquad(3.1)$$

的形状，可以用迭代法求解。

例　解联立方程

$$\left.\begin{aligned} 7x+y+2z&=10,\\ x+8y+3z&=8,\\ 2x+3y+9z&=6. \end{aligned}\right\}\qquad(3.2)$$

改写成

$$\left.\begin{aligned} x&=(10-y-2z)/7,\\ y&=(8-x-3z)/8,\\ z&=(6-2x-3y)/9. \end{aligned}\right\}\qquad(3.3)$$

取 $x_0=y_0=z_0=0$ 为第0次近似，而后按

$$\left.\begin{aligned} x_{k+1}&=(10-y_k-2z_k)/7,\\ y_{k+1}&=(8-x_k-3z_k)/8,\\ z_{k+1}&=(6-2x_k-3y_k)/9. \end{aligned}\right\}\quad(k=0,1,2,\cdots)\qquad(3.4)$$

逐次求第1次近似，第2次近似……，得到如表3.1所示的結果。

[注1]　如果在計算过程中每到小数第3位卽作舍入，最后不停地发生微小的振动現象。虽然如此，在本例中第10次近似已經給出可以說是相当滿意的解。精确的解是

$$x=132/103\approx1.28155,\ y=82/103\approx0.79612,\ z=12/103\approx0.11650.$$

[注2]　在本例中，每一次的修正量，对于一切变量来說，其大小每次約減小一半，而符号是正負相間的。这种有趣的現象在迭代法中是常常出現的(参看 p.17)。

代替公式(3.4)，如果在右边代入剛才得到的近似值，在大多

表 3.1 逐次近似解

k	x_k	y_k	z_k
0	0	0	0
1	1.429	1.000	0.667
2	1.095	0.571	0.016
3	1.342	0.857	0.233
4	1.240	0.745	0.083
5	1.298	0.814	0.143
6	1.271	0.784	0.107
7	1.286	0.801	0.123
8	1.279	0.793	0.114
9	1.283	0.797	0.118
10	1.281	0.795	0.116
11	1.282	0.796	0.117
12	1.281	0.796	0.116
13	1.282	0.796	0.117
14	1.281	0.796	0.116

数情形下可以多少加速收斂的过程[1]。此时,公式变成

$$\left.\begin{aligned} x_{k+1} &= (10-y_k-2z_k)/7, \\ y_{k+1} &= (8-x_{k+1}-3z_k)/8, \\ z_{k+1} &= (6-2x_{k+1}-3y_{k+1})/9, \end{aligned}\right\} \quad (3.5)$$

而所得的結果如表 3.2 所示。这方法通常叫做 **Gauss-Seidel** 迭代法[2]。

表 3.2 Gauss-Seidel 迭代法

k	x_k	y_k	z_k
0	0	0	0
1	1.429	0.821	0.075
2	1.290	0.811	0.110
3	1.281	0.799	0.116
4	1.281	0.796	0.117
5	1.281	0.796	0.117

由表 3.2 可以看到,在这种情形,第 4 次近似已經收斂了(以

[1] 不单如此,在电子計算机中,例如得到了 x_{k+1} 后,立即就可以用它代换 x_k 而把 x_k 去掉,因此可以节省記忆設备。

[2] 实际上,Gauss 絲毫也沒有涉及这一方法,Seidel 虽然提到了这一方法,但据說他认为最好不要用这种方法[参看在 19 頁引用的 Forsyth 的文章 (p.302)]。

后只是同一值的反复)，而且趋近这个值的方式也和表 3.1 大有不同(参看 p. 18)。

为了理解迭代法中逐次近似值的变动方式，差分方程[1]的理論是有用的。例如，首先考虑按公式(3.4)的迭代法。(3.4) 是关于 x_k, y_k, z_k 的联立差分方程。这种一切系数都是常数的綫性差分方程組的一般解可以在一组确定的解 $x_k=\xi$, $y_k=\eta$, $z_k=\zeta$ 上，加上由給定的方程組去掉常数项而得的齐次方程組的一般解("补余解")而得到。而且，补余解是形如

$$x_k=X\lambda^k,\quad y_k=Y\lambda^k,\quad z_k=Z\lambda^k \tag{3.6}$$

的项的和。确定的解 ξ, η, ζ 就是原方程組 (3.2)(即(3.3)) 的精确解。因此，在这种情形，"补余解"就是"誤差"。为了确定它的組成，将(3.6)代入由(3.4)去掉常数项而得的方程組中并加以整理，就可以看到 λ 和 X, Y, Z 必須满足

$$\left.\begin{aligned}\lambda X&=-(Y+2Z)/7,\\ \lambda Y&=-(X+3Z)/8,\\ \lambda Z&=-(2X+3Y)/9.\end{aligned}\right\} \tag{3.7}$$

为了得到"平凡解" $X=Y=Z=0$ 以外的解，X, Y, Z 的系数行列式必須为 0. 即

$$\begin{vmatrix}\lambda & 1/7 & 2/7\\ 1/8 & \lambda & 3/8\\ 2/9 & 3/9 & \lambda\end{vmatrix}=0. \tag{3.8}$$

展开上式左边的行列式得到

$$\lambda^3-\frac{13}{63}\lambda+\frac{1}{42}=0, \tag{3.9}$$

它的3根(特征值)是 $\lambda_1=-0.504$, $\lambda_2=0.379$, $\lambda_3=0.125$ (参看§5例2). 对于这三个根的每一个，$X:Y:Z$(特征向量的方向)是

---

[1] 参看本丛书福田著《差分方程》。

确定的。但是，每当 k 的值增加 1 时，(3.6) 就 λ 倍起来，因此，$|\lambda|$ 小的項迅速減小，而 $|\lambda|$ 大的項則減小得很慢(如果假定有 $|\lambda|>1$ 的項，那么这种項就逐漸增大)。在目前的情形对应于

$$\lambda_1=-0.504$$

的項減小得最慢。幷且，只有这个特征值是負数，它的影响以

$$+-+-\cdots$$

的振动形式出現。再加上以 $\lambda_2=0.379$ 和 $\lambda_3=0.125$ 的比率由一方面单調減小其幅度的阻尼运动，观察作为他們的总和的誤差的变化(图 3.1)，就可以对它有充分的理解。

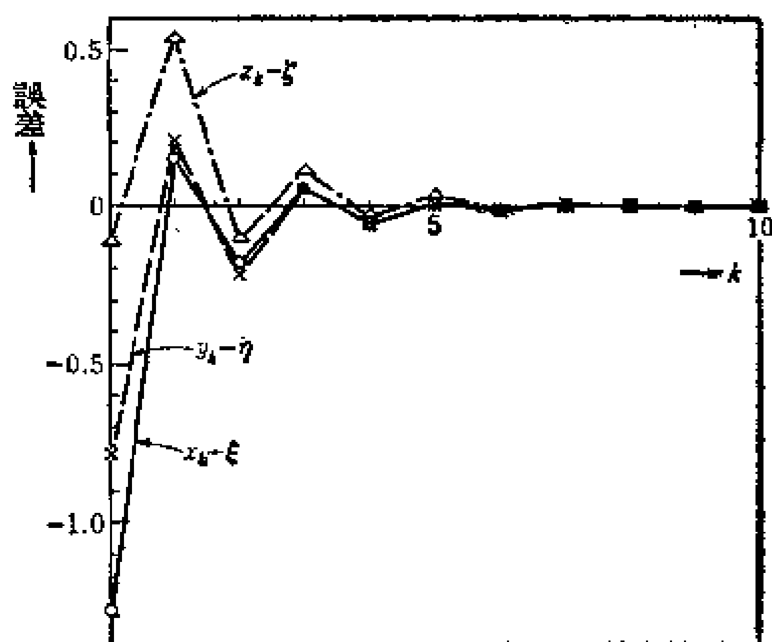


图 3.1 誤差的变化

其次，考虑按照(3.5)的迭代法(Gauss-Seidel 方法)中誤差的性状。对应于(3.8)，(3.9)的是

$$\begin{vmatrix} \lambda & 1/7 & 2/7 \\ \lambda/8 & \lambda & 3/8 \\ 2\lambda/9 & 3\lambda/9 & \lambda \end{vmatrix}=\lambda^3-\frac{7}{36}\lambda^2+\frac{1}{84}\lambda=0, \qquad (3.10)$$

它的根可以写成 λ_1, $\lambda_2=qe^{\pm i\theta}$, $\lambda_3=0$, 其中

$$q=\sqrt{\frac{1}{84}}=0.1091,\quad \cos\theta=\left(\frac{7}{36}\right)\frac{1}{2q}=0.8911,\quad \theta=0.471.$$

由此看到，在这种情形誤差的变化，除了在一瞬間即消失的部分($\lambda_3=0$)外,还有当 k 的值每增加 1 时振幅約減少成为 10 分之 1 的阻尼振动(周期为 $2\pi/\theta=13.3$)。

[注 1] 按照(3.5)得到的逐次近似值較按(3.4)得到的更快地收斂的原因在于，在(3.4)的情形，特征值的絕对值最大的約为 0.5，而在(3.5)的情形,特征值的絕对值最大的約为 0.1. 一般說来，在 3 元联立一次方程中,如果主对角綫上的系数都是 1，而其他系数的大小大約都等于 ε，那么,可以証明，在相当于(3.4)的情形，特征值的絕对值最大的也約等于 ε，而在相当于(3.5)的情形，約等于 $\varepsilon^{\frac{3}{2}}$. 但在元数更多的一般情形，目前还沒有确定的結論。

[注 2] 在表 3.1 中 $k\geqslant 11$ 的部分所看到的振动,与上面所說的不同，是因舍入的影响而引起的。为了說明这一問題,作变换

$$x=1.281+0.001u,\quad y=0.796+0.001v,\quad z=0.116+0.001w,$$

关于 u, v, w 和(3.4)对应的关系式是

$$u_{k+1}=(5-v_k-2w_k)/7,\quad v_{k+1}=(3-u_k-3w_k)/8,$$
$$w_{k+1}=(6-2u_k-3v_k)/9.$$

而且,因为 u, v, w 經过舍入都将成为整数,因此,如令 $u_k=v_k=w_k=0$ ($k=12$ 或 14 等)就得到

$$u_{k+1}=\frac{5}{7}\approx 1,\quad v_{k+1}=\frac{3}{8}\approx 0,\quad w_{k+1}=\frac{6}{9}\approx 1^{❶},$$

如令 $u_k=1,\quad v_k=0,\quad w_k=1,$

就得到 $u_{k+1}=3/7\approx 0,\quad v_{k+1}=-1/8\approx 0,\quad w_{k+1}=4/9\approx 0,$

因而(0, 0, 0)和(1, 0, 1)交替地出現。在 Gauss-Seidel 方法中也可能发生同样的情形。

§4 共軛斜量法

在联立一次方程的解法中最近出現了所謂共軛斜量法

---

❶ $x\approx x_0$ 是表示 x 經过舍入后等于 x_0 的記号。

(method of conjugate gradients 簡記为 cg 法) 的有趣的新方法[1]。与迭代法相同,这也是一个逐次逼近精确解的方法,但在經过有限次計算 (在 n 元的情形如果沒有舍入的話,是 n 次) 即告結束的这一点上是和消去法相同的。而且,由于它的程序或者成为簡单手續的反复,或者需要的記忆容量較少,或者在 0 系数很多的情形能够将此性质比較长久地保存下去等原因,因此被认作是特别适用于快速电子計算机的方法。

用 $\boldsymbol{x}$ 表示未知数的向量,A 表示系数矩陣,$\boldsymbol{k}$ 表示右边的向量,而把方程組写成

$$A\boldsymbol{x}=\boldsymbol{k} \tag{4.1}$$

的形状。为簡单計以下討論 A 是对称矩陣而且具有正定符号的情形。

一般地用

$$\boldsymbol{r}_i=\boldsymbol{k}-A\boldsymbol{x}_i \tag{4.2}$$

定义第 i 近似解 $\boldsymbol{x}_i$ 的殘差 ($i=0, 1, 2, \cdots$)。設精确解为 $\boldsymbol{h}$,那么,由于 $A\boldsymbol{h}=\boldsymbol{k}$,因此,也可以写成 $\boldsymbol{r}_i=A(\boldsymbol{h}-\boldsymbol{x}_i)$,由此推出,

$$\begin{aligned} f(\boldsymbol{x}_i) &\equiv (\boldsymbol{h}-\boldsymbol{x}_i,\ A(\boldsymbol{h}-\boldsymbol{x}_i)) = (A^{-1}\boldsymbol{r}_i,\ \boldsymbol{r}_i) \\ &= (\boldsymbol{h}-\boldsymbol{x}_i,\ \boldsymbol{r}_i). \end{aligned} \tag{4.3}$$

$f(\boldsymbol{x}_i)$ 叫做誤差函数。它把精确解 $\boldsymbol{h}$ 与近似解 $\boldsymbol{x}_i$ 的差 (向量) 的"大小"以矩陣 A 为度量張量而表示了出来,也就是把殘差向量 $\boldsymbol{r}_i$ 的"大小"用逆矩陣 A^{-1} 作了度量。对于精确解 $\boldsymbol{h}$,$f(\boldsymbol{h})=0$,而且反之,如果把誤差函数逐漸縮小,最后終于到达了 $f(\boldsymbol{x}_m)=0$ 的話,这 $\boldsymbol{x}_m$ 和 $\boldsymbol{h}$ 是全等的。

假設对于第 i 近似解 $\boldsymbol{x}_i$,加上在某一向量 $\boldsymbol{p}_i$ 方向上的修正

---

[1] M. R. Hestenes and E. Stiefel, J. Res. Nat. Bur. Standards, 49 (1952); paper 2379; G. E. Forsyth, Bull. Amer. Math. Soc., 59(1953) 299～329; M. R. Hestenes, Proc. 6th. Symp. Appl. Math. (1956), 83～102.

而得到第$(i+1)$近似解:

$$\boldsymbol{x}_{i+1}=\boldsymbol{x}_i+\alpha \boldsymbol{p}_i. \tag{4.4}$$

此时,为了使$f(\boldsymbol{x}_{i+1})$达到最小,取

$$\alpha=\alpha_i\equiv(\boldsymbol{p}_i,\boldsymbol{r}_i)/(\boldsymbol{p}_i,A\boldsymbol{p}_i) \tag{4.5}$$

就可以了。

証明　$\boldsymbol{x}_{i+1}$的殘差是$\boldsymbol{r}_{i+1}=\boldsymbol{k}-A(\boldsymbol{x}_i+\alpha\boldsymbol{p}_i)=\boldsymbol{k}-A\boldsymbol{x}_i-\alpha A\boldsymbol{p}_i=\boldsymbol{r}_i-\alpha A\boldsymbol{p}_i$,由此推出,$f(\boldsymbol{x}_{i+1})=(A^{-1}\boldsymbol{r}_{i+1},\boldsymbol{r}_{i+1})=(A^{-1}\boldsymbol{r}_i-\alpha\boldsymbol{p}_i,\boldsymbol{r}_i-\alpha A\boldsymbol{p}_i)=(A^{-1}\boldsymbol{r}_i,\boldsymbol{r}_i)-2\alpha(\boldsymbol{p}_i,\boldsymbol{r}_i)+\alpha^2(\boldsymbol{p}_i,A\boldsymbol{p}_i)=(A^{-1}\boldsymbol{r}_i,\boldsymbol{r}_i)-(\boldsymbol{p}_i,\boldsymbol{r}_i)^2/(\boldsymbol{p}_i,A\boldsymbol{p}_i)+(\boldsymbol{p}_i,A\boldsymbol{p}_i)\{(\boldsymbol{p}_i,\boldsymbol{r}_i)/(\boldsymbol{p}_i,A\boldsymbol{p}_i)-\alpha\}^2$.　　証毕

这样一来,

$$\boldsymbol{x}_{i+1}=\boldsymbol{x}_i+\alpha_i\boldsymbol{p}_i, \tag{4.6}$$

$$\boldsymbol{r}_{i+1}=\boldsymbol{r}_i-\alpha_i A\boldsymbol{p}_i, \tag{4.7}$$

因而

$$f(\boldsymbol{x}_{i+1})=f(\boldsymbol{x}_i)-(\boldsymbol{p}_i,\boldsymbol{r}_i)^2/(\boldsymbol{p}_i,A\boldsymbol{p}_i). \tag{4.8}$$

而且,新的殘差向量$\boldsymbol{r}_{i+1}$和$\boldsymbol{p}_i$正交,即

$$(\boldsymbol{p}_i,\boldsymbol{r}_{i+1})=0. \tag{4.9}$$

証明　$(\boldsymbol{p}_i,\boldsymbol{r}_{i+1})=(\boldsymbol{p}_i,\boldsymbol{r}_i-\alpha_i A\boldsymbol{p}_i)=(\boldsymbol{p}_i,\boldsymbol{r}_i)-\alpha_i(\boldsymbol{p}_i,A\boldsymbol{p}_i)=0$.

但是,决定修正方向的向量$\boldsymbol{p}_i$可以有种种不同的取法。在共軛斜量法中,是在到此为止的一切殘差向量$\boldsymbol{r}_0,\boldsymbol{r}_1,\cdots,\boldsymbol{r}_i$所張成的空間中选取$\boldsymbol{p}_i$,而且使对于新的近似值,誤差函数取尽可能小的值。换句話說,設第$i+1$近似值是

$$\boldsymbol{x}_{i+1}=\boldsymbol{x}_i+\alpha_i\boldsymbol{r}_i+\alpha_{i-1}\boldsymbol{r}_{i-1}+\cdots+\alpha_0\boldsymbol{r}_0, \tag{4.10}$$

而后决定系数$\alpha_i,\alpha_{i-1},\cdots,\alpha_0$,使$f(\boldsymbol{x}_{i+1})$达到最小❶。此时,新的殘差是

$$\begin{aligned}\boldsymbol{r}_{i+1}&=\boldsymbol{k}-A(\boldsymbol{x}_i+\alpha_i\boldsymbol{r}_i+\alpha_{i-1}\boldsymbol{r}_{i-1}+\cdots+\alpha_0\boldsymbol{r}_0)\\&=\boldsymbol{r}_i-(\alpha_i A\boldsymbol{r}_i+\alpha_{i-1}A\boldsymbol{r}_{i-1}+\cdots+\alpha_0 A\boldsymbol{r}_0).\end{aligned} \tag{4.11}$$

---

❶ 这里的系数应写成$\alpha_i^i,\alpha_{i-1}^i,\cdots,\alpha_0^i$比較恰当。由于过份繁杂,因此省去上面的肩号$i$.

因此

$$f(\boldsymbol{x}_{i+1})=(A^{-1}\boldsymbol{r}_{i+1},\ \boldsymbol{r}_{i+1})=(A^{-1}\boldsymbol{r}_i-\sum_{j=0}^{i}\alpha_j\boldsymbol{r}_j,\boldsymbol{r}_i-\sum_{k=0}^{i}\alpha_kA\boldsymbol{r}_k)$$

$$=(A^{-1}\boldsymbol{r}_i,\ \boldsymbol{r}_i)-2\sum_{j=0}^{i}\alpha_j(\boldsymbol{r}_j,\ \boldsymbol{r}_i)+\sum_{j=0}^{i}\sum_{k=0}^{i}\alpha_j\alpha_k(\boldsymbol{r}_j,\ A\boldsymbol{r}_k),\tag{4.12}$$

为了使它成为极小，只要选取 $\alpha_0,\ \alpha_1,\ \cdots,\ \alpha_i$，使它們满足

$$\frac{\partial f(\boldsymbol{x}_{i+1})}{\partial\alpha_j}=-2(\boldsymbol{r}_j,\ \boldsymbol{r}_i)+2\sum_{k=0}^{i}\alpha_k(\boldsymbol{r}_j,\ A\boldsymbol{r}_k)=0\quad(j=0,\ \cdots,\ i)\tag{4.13}$$

就可以了[❶]。利用(4.11)还可以把(4.13)写成(略去系数 -2)

$$(\boldsymbol{r}_j,\ \boldsymbol{r}_{i+1})=0\quad(j=0,\ \cdots,\ i).\tag{4.14}$$

换句話說，按照上述的方法确定第 $(i+1)$ 近似值，新的殘差就和到此为止的一切殘差正交。由此推出，在共軛斜量法中，关系式

$$(\boldsymbol{r}_i,\ \boldsymbol{r}_j)=0\quad(i\neq j)\tag{4.15}$$

一般地成立。共軛斜量法可以这样的**殘差的正交性**作为其特征。

现在假定，由 $\boldsymbol{x}_i$ 轉換到 $\boldsymbol{x}_{i+1}$ 时修正的向量为 $\boldsymbol{q}_i$，那么

$$\boldsymbol{q}_i=\alpha_i\boldsymbol{r}_i+\alpha_{i-1}\boldsymbol{r}_{i-1}+\cdots+\alpha_0\boldsymbol{r}_0,\tag{4.16}$$

$$\boldsymbol{r}_{i+1}=\boldsymbol{r}_i-A\boldsymbol{q}_i,\qquad\therefore\ A\boldsymbol{q}_i=\boldsymbol{r}_i-\boldsymbol{r}_{i+1}.\tag{4.17}$$

于是，利用殘差的正交性(4.15)得到

$$(\boldsymbol{r}_j,\ A\boldsymbol{q}_i)=\begin{cases}0 & (j<i),\\ |\boldsymbol{r}_i|^2 & (j=i),\\ -|\boldsymbol{r}_{i+1}|^2 & (j=i+1),\\ 0 & (j>i+1).\end{cases}\tag{4.18}$$

特别值得注意的是，$A\boldsymbol{q}_i$ 和 $\boldsymbol{r}_0,\ \cdots,\ \boldsymbol{r}_{i-1}$ 正交。但是，由于一般来說，$\boldsymbol{q}_j$ 是 $\boldsymbol{r}_0,\ \cdots,\ \boldsymbol{r}_j$ 的一次式(綫性組合)，由此可以看到，$j<i$ 时，$(\boldsymbol{q}_j,\ A\boldsymbol{q}_i)=0$. 因为 A 是对称矩陣，因而 $(\boldsymbol{q}_i,\ A\boldsymbol{q}_j)=(\boldsymbol{q}_j,\ A\boldsymbol{q}_i)$，

❶ 具体地确定 $\alpha_0,\ \cdots,\ \alpha_i$ 的問題，以后将用更为間接的方法导出其計算式。

由此推出，一般地

$$(\boldsymbol{q}_i,\ A\boldsymbol{q}_j)=0 \quad (i\neq j) \tag{4.19}$$

成立。换句話說，每一次的修正向量 $\boldsymbol{q}_0$，$\boldsymbol{q}_1$，$\boldsymbol{q}_2$，… 是彼此共軛(A-正交)的。在这种意义下，共軛斜量法是共軛方向法(method of conjugate direction，簡記为 cd 法)的一种。

其次，把(4.17)中的 i 换成 j 而代入(4.19)，就得到

$$(\boldsymbol{q}_i,\ \boldsymbol{r}_j)=(\boldsymbol{q}_i,\ \boldsymbol{r}_{j+1}) \quad (i\neq j).$$

由此推出

$$(\boldsymbol{q}_i,\ \boldsymbol{r}_0)=(\boldsymbol{q}_i,\ \boldsymbol{r}_1)=\cdots=(\boldsymbol{q}_i,\ \boldsymbol{r}_{i-1})=(\boldsymbol{q}_i,\ \boldsymbol{r}_i). \tag{4.20}$$

在此代入 $\boldsymbol{q}_i$ 的表示式(4.16)并利用殘差的正交性(4.15)，就得到

$$\alpha_0|\boldsymbol{r}_0|^2=\alpha_1|\boldsymbol{r}_1|^2=\cdots=\alpha_{i-1}|\boldsymbol{r}_{i-1}|^2=\alpha_i|\boldsymbol{r}_i|^2.$$

由此可以把(4.16)改写成

$$\boldsymbol{q}_i=\alpha_i\left(\boldsymbol{r}_i+\frac{|\boldsymbol{r}_i|^2}{|\boldsymbol{r}_{i-1}|^2}\boldsymbol{r}_{i-1}+\cdots+\frac{|\boldsymbol{r}_i|^2}{|\boldsymbol{r}_0|^2}\boldsymbol{r}_0\right). \tag{4.21}$$

为便利計，定义方向向量为

$$\boldsymbol{p}_i=\boldsymbol{r}_i+\frac{|\boldsymbol{r}_i|^2}{|\boldsymbol{r}_{i-1}|^2}\boldsymbol{r}_{i-1}+\cdots+\frac{|\boldsymbol{r}_i|^2}{|\boldsymbol{r}_0|^2}\boldsymbol{r}_0, \tag{4.22}$$

关于这些方向向量有下列的递推公式成立：

$$\boldsymbol{p}_i=\boldsymbol{r}_i+\frac{|\boldsymbol{r}_i|^2}{|\boldsymbol{r}_{i-1}|^2}\boldsymbol{p}_{i-1} \quad (i\geqslant 1), \quad \boldsymbol{p}_0=\boldsymbol{r}_0. \tag{4.23}$$

修正向量 $\boldsymbol{q}_i=\alpha_i\boldsymbol{p}_i$ 和方向向量 $\boldsymbol{p}_i$ 有相同的方向。今后把系数 α_i 記作 a_i (按照第 20 頁脚注的記号，这就表示把 α_i^i 記作 a_i)。只要 $|\boldsymbol{r}_i|\neq 0$，就有 $a_i\neq 0$。这是因为，假如 $a_i=0$，那么就得 $\boldsymbol{q}_i=0$，由(4.17)，$\boldsymbol{r}_i=\boldsymbol{r}_{i+1}$，由此推出，$|\boldsymbol{r}_i|^2=(\boldsymbol{r}_i,\ \boldsymbol{r}_{i+1})=0$。以下在 $|\boldsymbol{r}_i|\neq 0$ 的条件下进行討論。

由(4.22)和殘差的正交性(4.15)可以看到

$$(\boldsymbol{p}_i,\ \boldsymbol{r}_j)=\begin{cases}|\boldsymbol{r}_i|^2 & (j\leqslant i),\\ 0 & (j>i).\end{cases} \tag{4.24}$$

用 $a_i \boldsymbol{p}_i$ 代(4.18)中的 $\boldsymbol{q}_i$ 还可以得到

$$(\boldsymbol{r}_j, A\boldsymbol{p}_i) = \begin{cases} |\boldsymbol{r}_i|^2/a_i & (j=i), \\ -|\boldsymbol{r}_{i+1}|^2/a_i & (j=i+1), \\ 0 & (\text{在其余的情形}). \end{cases} \tag{4.25}$$

特别是，利用$(\boldsymbol{r}_j, A\boldsymbol{p}_i)=0(j<i)$，由(4.22)可以推出

$$(\boldsymbol{r}_i, A\boldsymbol{p}_i) = (\boldsymbol{p}_i, A\boldsymbol{p}_i). \tag{4.26}$$

由此还得到$(\boldsymbol{p}_i, A\boldsymbol{p}_i) = |\boldsymbol{r}_i|^2/a_i$，也就是

$$a_i = \frac{|\boldsymbol{r}_i|^2}{(\boldsymbol{p}_i, A\boldsymbol{p}_i)}. \tag{4.27}$$

这一表示式用于在实际計算中求 a_i。

以上所述的共軛斜量法的程序可以归結如下：

开始的程序 任意取定第 0 近似解 $\boldsymbol{x}_0$，利用下面的式子計算殘差 $\boldsymbol{r}_0$，$c_0=|\boldsymbol{r}_0|^2$，以及方向向量 $\boldsymbol{p}_0$：

$$\boldsymbol{r}_0=\boldsymbol{k}-A\boldsymbol{x}_0, \quad c_0=|\boldsymbol{r}_0|^2, \quad \boldsymbol{p}_0=\boldsymbol{r}_0.$$

一般的手續 假設已經求得第 i 近似值 $\boldsymbol{x}_i$，殘差

$$\boldsymbol{r}_i, \quad c_i=|\boldsymbol{r}_i|^2,$$

以及方向向量 $\boldsymbol{p}_i$，求 $A\boldsymbol{p}_i$ 而且依次計算：

$$d_i=(\boldsymbol{p}_i, A\boldsymbol{p}_i), \ a_i=c_i/d_i, \ \boldsymbol{x}_{i+1}=\boldsymbol{x}_i+a_i\boldsymbol{p}_i,$$

$$\boldsymbol{r}_{i+1}=\boldsymbol{r}_i-a_iA\boldsymbol{p}_i, \ c_{i+1}=|\boldsymbol{r}_{i+1}|^2,$$

$$b_i=c_{i+1}/c_i, \ \boldsymbol{p}_{i+1}=\boldsymbol{r}_{i+1}+b_i\boldsymbol{p}_i.$$

这样就可以确定出逐次近似解 $\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \cdots$；殘差 $\boldsymbol{r}_0, \boldsymbol{r}_1, \boldsymbol{r}_2, \cdots$；方向向量 $\boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{p}_2, \cdots$。这些殘差具有正交性(4.10)，因此，如果沒有舍入誤差，那么在 n 維的情形，一定有 $m\leqslant n$ 的某个 m 存在，使 $\boldsymbol{r}_m=0$(在 n 維空間中，由多于 n 个的向量組成的正交向量組是不存在的)。由此推出 $\boldsymbol{x}_m=\boldsymbol{h}$(精确解)，即至多在 n 次以后就到达了解的精确值。

例 在§3的例中，

$$A=\begin{pmatrix}7&1&2\\1&8&3\\2&3&9\end{pmatrix},\quad k=\begin{pmatrix}10\\8\\6\end{pmatrix}.$$

以下設第0近似解为 $x_0=(0, 0, 0)$ 而試用共軛斜量法来求解。

計算的紀录，例如可以写成表4.1的形式。

[注1] 如果用表4.1的 x_3 直接計算殘差 $r_3=k-Ax_3$，就得到

$$r_3=0.00004,\ 0.00007,\ 0.00007.$$

表4.1中的 r_3 所以和这些值不同是由于舍入誤差的影响。

[注2] 为了計算过程中的驗算，每一次得到新的 p_i 时应該驗証一下是否 $(Ap_{i-1}, p_i)=0$. 在表4.1中，

$$(Ap_0, p_1)=-0.00204,\quad (Ap_1, p_2)=0.00005,$$

这种程度的舍入誤差是可事前估計到(注意 Ap_0 的各分量都在100左右)。

在共軛斜量法中，精确解 h 和近似解 x_i 的差向量的长度

$$|h-x_i|$$

是随着 i 的增大而单調減小的。

証明 設第一个成为0的殘差是 r_m，那么，$x_m=h$，因此，对于任意的 $i(\leqslant m-1)$，

$$h=x_i+a_ip_i+a_{i+1}p_{i+1}+\cdots+a_{m-1}p_{m-1}. \tag{4.28}$$

因而
$$h-x_i=a_ip_i+a_{i+1}p_{i+1}+\cdots+a_{m-1}p_{m-1},$$

而且❶
$$h-x_{i+1}=a_{i+1}p_{i+1}+\cdots+a_{m-1}p_{m-1}.$$

由此推出

$$\begin{aligned}|h-x_i|^2&=|h-x_{i+1}+a_ip_i|^2\\&=|h-x_{i+1}|^2+a_i^2|p_i|^2+2a_i(h-x_{i+1}, p_i).\end{aligned} \tag{4.29}$$

但由于

$$\begin{aligned}(h-x_{i+1}, p_i)&=(a_{i+1}p_{i+1}+\cdots+a_{m-1}p_{m-1}, p_i)\\&=\sum_{j=i+1}^{m-1}a_j(p_j, p_i),\end{aligned} \tag{4.30}$$

而且由(4.22)和(4.15)，当 $j>i$ 时，

$$(p_j, p_i)=|r_j|^2+\frac{|r_j|^2|r_i|^2}{|r_{i-1}|^2}+\cdots+\frac{|r_j|^2|r_i|^2}{|r_0|^2}=|r_j|^2\frac{|p_i|^2}{|r_i|^2}, \tag{4.31}$$

---

❶ $i=m-1$ 时命題是显見的，因而以下在 $i<m-1$ 的假定下进行討論。

表4.1 共轭斜量法

阶段	向量	向量的分量			辅助系数	説明
0	x_0	0	0	0		
	r_0	10	8	6	$c_0=200$	$r_0=k-Ax_0,\ p_0=r_0,$
	p_0	10	8	6	$d_0=2224$	$c_0=\|r_0\|^2,\ d_0=(p_0,\ Ap_0),$
	Ap_0	90	92	98	$a_0-0.08992\ 81$	$a_0=c_0/d_0$
1	x_1	0.89928	0.71942	0.53957	$c_1=11.62206$	$x_1=x_0+a_0p_0,\ r_1=r_0-a_0Ap_0,$
	r_1	1.90647	−0.27339	−2.81295	$b_0=0.05811\ 03$	$c_1=\|r_1\|^2,\ b_0=c_1/c_0,$
	p_1	2.48757	0.19149	−2.46429	$d_1=71.86490$	$p_1=r_1+b_0p_0,$
	Ap_1	12.67590	−3.37338	−16.62900	$a_1=0.16172\ 10$	$d_1=(p_1,\ Ap_1),\ a_1=c_1/d_1$
2	x_2	1.30157	0.75039	0.14104	$c_2=0.10995\ 97$	$x_2=x_1+a_1p_1,\ r_2=r_1-a_1Ap_1,$
	r_2	−0.14349	0.27216	−0.12369	$b_1=0.00946\ 13$	$c_2=\|r_2\|^2,\ b_1=c_2/c_1,$
	p_2	−0.11995	0.27397	−0.14701	$d_2=0.65885\ 20$	$p_2=r_2+b_1p_1,$
	Ap_2	−0.85970	1.63078	−0.74108	$a_2=0.16689\ 59$	$d_2=(p_2,\ Ap_2),\ a_2=c_2/d_2$
3	x_3	1.28155	0.79611	0.11650		$x_3=x_2+a_2p_2,$
	r_3	−0.00001	−0.00001	−0.00001		$r_3=r_2-a_2Ap_2$

因此，最后得到

$$|h-x_i|^2-|h-x_{i+1}|^2=a_i^2|p_i|^2+2a_i\sum_{j=i+1}^{m-1}a_j|r_j|^2\frac{|p_i|^2}{|r_i|^2}. \qquad (4.32)$$

如由(4.27)看到的那样，$a_i, a_{i+1}, \cdots, a_{m-1}$都是正数，因此，(4.32)的右边是正的。也就是說，$|h-x_i|^2>|h-x_{i+1}|^2$，由此推出$|h-x_i|>|h-x_{i+1}|$.

証毕

可以說，在如上所述的那样誤差逐漸減小的点上，共軛斜量法具有逐次逼近法的性质。

以上是在系数矩陣A为对称的而且具有正定符号的条件下进行討論的，如果A不是对称的，用它的轉置矩陣A'左乘(4.1)得到

$$A'Ax=A'k, \qquad (4.33)$$

因为$A'A$是对称的，因此可以适用上述的方法。不过，这样的做法在理論上与下述方式是等价的：

开始的程序 任意取定第0近似解x_0而計算

$$r_0=k-Ax_0, \quad p_0=A'r_0, \quad a_0=|A'r_0|^2/|Ap_0|^2.$$

一般的手續 在x_i, r_i, p_i, a_i已知时順次計算：

$$x_{i+1}=x_i+a_ip_i,\ r_{i+1}=r_i-a_iAp_i,\ b_i=|A'r_{i+1}|^2/|A'r_i|^2,$$

$$p_{i+1}=A'r_{i+1}+b_ip_i,\ a_{i+1}=|A'r_{i+1}|^2/|Ap_{i+1}|^2.$$

这一方式在数值計算上一般說来較先导出(4.33)而后使用前述的方法为优越（注意在后面的方式中不需要計算矩陣的乘积AA'）。

最后，在矩陣A非正規的情形，或者(i)在計算过程中得到$r_m=0$而到达所求的解，或者(ii)产生$r_m\neq 0$而$Ap_m=0$的情形，解法不能継續进行。在(ii)的情形給定的方程組是不相容的（在这种情形，也可以由到此为止的計算得到最小二乘解）❶。

---

❶ 关于这一点，詳見19頁上所引的文献。

§5 高次代数方程

一元二次方程有解的公式，因此，平常就可以用它来求数值解，在 3 次以上的情形一般不能这样做。虽然在 3 次和 4 次的情形，也有所謂代数解法，但对于数值計算来說是不很便利的。更不用說在 5 次以上的情形，一般来說，就沒有代数的解法。因此，在本节中将討論 3 次以上方程的数值解法。

例 1 求方程

$$f(x)=126x^3-26x+3=0 \tag{5.1}$$

的 3 个根(这方程和方程(3.9)是等价的)。

在解这类問題时，首先应該作出 $f(x)$ 的大致的图形。对于 $x=0$, ±1, ±2 等簡单的值求出 $f(x)$，就可以預測，实根可能在 -1 和 1 之間。于是，对于 $x=\pm0.1$, ±0.2, ±0.3, …，計算 $f(x)$ 而画出图形，如图 5.1 所示，由此看到，在区間

$$(-0.6,\ -0.5),\ (0.1,\ 0.2),\ (0.3,\ 0.4)$$

內各有一个实根。

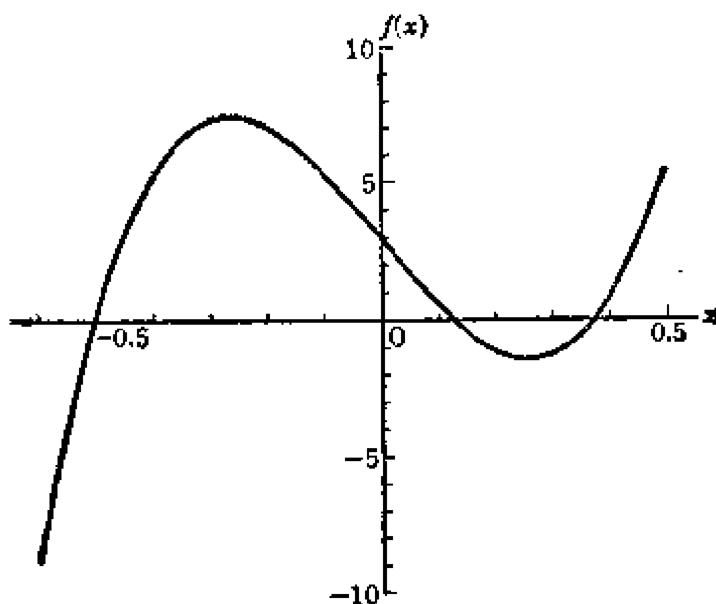


图 5.1 $f(x)$ 的图形

[注] 在作这样的計算时，一般来說，用綜合除法是便利的。設多項式

$$f(x)=a_0x^3+a_1x^2+a_2x+a_3 \tag{5.2}$$

用$(x-\alpha)$除得的商是

$$g(x)=b_0x^2+b_1x+b_2 \tag{5.3}$$

而剩余是 R，那么，恒等式

$$f(x)=(x-\alpha)g(x)+R \tag{5.4}$$

成立。代入(5.2)和(5.3)并比較 x 的各次幂的系数得到

$$a_0=b_0,\ a_1=b_1-\alpha b_0,\ a_2=b_2-\alpha b_1,\ a_3=R-\alpha b_2. \tag{5.5}$$

由此推出

$$b_0=a_0,\ b_1=a_1+\alpha b_0,\ b_2=a_2+\alpha b_1,\ R=a_3+\alpha b_2. \tag{5.6}$$

如用笔算进行，写成如下的方式是便利的：

a_0	a_1	a_2	a_3	α
	αb_0	αb_1	αb_2	
b_0	b_1	b_2	R	

例：

126	0	−26	3	0.2
	25.2	5.04	−4.192	
126	25.2	−20.96	−1.192	

用台式計算机时不必記录 αb_0，αb_1，αb_2，而按表 5.1 的方式記录更为恰当。

在(5.4)中令 $x=\alpha$ 就可以看到

$$R=f(\alpha) \tag{5.7}$$

(剩余定理)。这样就算出了 $f(\alpha)$ 的值。

表 5.1 綜合除法

	α
a_0	b_0
a_1	b_1
a_2	b_2
a_3	R

其次，在上面的例中，也可以考虑将已知有根的各区間更进一步細分而計算 $f(-0.51)$，$f(-0.52)$，… 的做法，但是，如果这样做的話，例如要将根求至小数第5位，就需要非常繁复的計算。

因此，以下将用效率更高的方法来求进一步的近似值。这种方法中一般說来最便利的是 Newton-Raphson 的迭代法。假設已經得到了 $f(x)=0$ 的根的近似值 x_i，設精确值为 x_i+q，那么应当有 $f(x_i+q)=0$。将此式的左边作 Taylor 展开得到

$$f(x_i)+qf'(x_i)+\frac{q^2}{2}f''(x_i)+\cdots=0. \tag{5.8}$$

如果 x_i 和根的精确值已經很接近，那么 q 的值很小，因此，它的平

方以上的項可以略去不計，而 q 的近似式可以解

$$f(x_i)+qf'(x_i)=0$$

而得到，即

$$q=-f(x_i)/f'(x_i).$$

因此，可以认为

$$x_{i+1}=x_i-f(x_i)/f'(x_i) \tag{5.9}$$

是比之 x_i 更接近于精确解的近似值。反复使用这种方法，直至 x_{i+1} 和 x_i 几乎沒有差别时就可以认作已經向精确值收斂了。

［注］ 在应用(5.9)时，不单 $f(x_i)$ 的計算可以利用綜合除法，$f'(x_i)$ 的計算也可以利用綜合除法。設(5.3)的 $g(x)$ 用 $x-\alpha$ 除得的商是

$$h(x)=c_0x+c_1,$$

剩余是 R'，那么，恒等式 $g(x)=(x-\alpha)h(x)+R'$ 成立。代入(5.4)得到

$$f(x)=(x-\alpha)^2h(x)+R'(x-\alpha)+R. \tag{5.10}$$

微分这一恒等式而令 $x=\alpha$ 就得到

$$f'(\alpha)=R'. \tag{5.11}$$

由 $g(x)$ 求 $h(x)$ 和 R' 的計算可以用和 (5.6) 同样的方法。将此計算附加在(5.6)的計算記录上或写成表 5.2 的形式是便利的。

$$\begin{array}{cccc|l} a_0 & a_1 & a_2 & a_3 & \alpha \\ & \alpha b_0 & \alpha b_1 & \alpha b_2 & \\ \hline b_0 & b_1 & b_2 & R=f(\alpha) & \\ & \alpha c_0 & \alpha c_1 & & \\ \hline c_0 & c_1 & R'=f'(\alpha) & & \end{array}$$

表 5.2 $f(\alpha)$，$f'(\alpha)$的計算

	α	
a_0	b_0	c_0
a_1	b_1	c_1
a_2	b_2	$R'=f'(\alpha)$
a_3	$R=f(\alpha)$	

例 2 求将方程(5.1)的 3 个根各确定到小数第 5 位。

首先，求接近于 -0.5 的根。假定用台式計算机进行計算，按照表 5.2 的形式进行記录，就得到表 5.3，由此看出 -0.50361 到小数第 5 位是正确的。

其次，类似地可以用表 5.4 和表 5.5 計算得其他 2 根为 0.37881 和 0.12481.

表 5.3　$f(x)=126x^3-26x+3=0$ 的根（接近于 -0.5 的根）

a_i	$x_0=-0.5$		$x_1=-0.50365$		$x_2=-0.50361$	
126	126	126	126	126	126	126
0	−63	−126	−63.45990	−126.91980	−63.45486	−126.90972
−26	5.5	63.5	5.96158	69.88474	5.95650	69.86950
3	0.25		−0.00255		+0.00025	
修　正	−0.00365		+0.00003 65		−0.00000 36	

［注1］ 在第2近似值 x_2 以后的地方，$f'(x_i)$ 的值几乎不变，因此，不一一重新計算也可以。

［注2］ 用表5.3求得第1根 -0.50361 时，同时求得了 $f(x)$ 用 $(x+0.50361)$ 除得的商 $g(x)=126x^2-63.45486x+5.95650$，因此，也可以解 $g(x)=0$ 而得到其他的两根：

$$x=\frac{31.72743\pm\sqrt{31.72743^2-126\times5.95650}}{126}$$

$$=\frac{31.72743\pm\sqrt{256.11081}}{126}$$

$$=\frac{31.72743\pm16.00346}{126}=\begin{cases}0.37882,\\0.12479.\end{cases}$$

但是，这方法不象表5.4和表5.5的計算那样地直接利用 $f(x)$，而利用了多少含有誤差的 $g(x)$，因此，最后一位的数值是不可靠的。为了防止这种現象，必須把第1个根計算到更多的位数，而把 $g(x)$ 取到更多的位数才可以。

用“代数解法”求解的情形，利用(5.9)反复进行計算有时較为方便。以下将举几个这样的例子。

例3　求平方根的 Newton 方法

設 $f(x)=x^2-a$，那么 $f(x)=0$ 的精确解是 $x=\pm\sqrt{a}$．对应地，(5.9)成为

$$x_{i+1}=x_i-\frac{x_i^2-a}{2x_i}=\frac{1}{2}\left(x_i+\frac{a}{x_i}\right). \tag{5.12}$$

表 5.4 $f(x)\equiv 126x^3-26x+3=0$ 的根(接近于 0.4 的根)

a_i	$x_0=0.4$		$x_1=0.38074$		$x_2=0.37883$		$x_3=0.37881$	
126	126	126	126	126	126	126	126	126
0	50.4	100.8	47.97324	95.94648	47.73258	95.46516	47.73006	95.46012
−26	−5.84	34.48	−7.73467	28.79599	−7.91747	28.24760	−7.91938	28.24187
3	0.664		0.05510		0.00062		0.00006	
修 正	−0.01926		−0.00191 36		−0.00002 19		−0.00000 21	

表 5.5 $f(x)\equiv 126x^3-26x+3=0$ 的根(接近于 0.1 的根)

a_i	$x_0=0.1$		$x_1=0.12367$		$x_2=0.12480$		$x_3=0.12481$	
126	126	126	126	126	126	126	126	126
0	12.6	25.2	15.58242	31.16484	15.72480	31.44960	15.72606	31.45212
−26	−24.74	−22.22	−24.07292	−20.21876	−24.03754	−20.11263	−24.03723	−20.11169
3	0.526		0.02290		0.00011 50		−0.00008 67	
修 正	+0.02867		+0.00113 26		+0.00000 57		−0.00000 43	

假設用某些方法得到了 $\sqrt{a}$ 的近似值 x_0，使用(5.12)容易提高它的精确度。例如用4位表❶求得 x_0 时，使用(5.12)一次就可以得到准确到第7位乃至第8位的值。在用电子計算机时❷，对于 $1<a<100$ 一律令 $x_0=3$ 而利用(5.12)反复进行計算，7次左右就可以得到8位左右的精确度。例如，对于 $a=99$，令 $x_0=3$，就得到

$$x_1=18,\ x_2=11.75,\ x_3=10.08776\,596,\ x_4=9.95081\,680,$$
$$x_5=9.94987\,442,\ x_6=9.94987\,437,\ x_7=9.94987\,437.$$

例4 倒数 設 $f(x)=1/x-a$，那么对应的(5.9)是

$$x_{i+1}=x_i-\left(\frac{1}{x_i}-a\right)\Big/\left(\frac{-1}{x_i^2}\right)=x_i+(x_i-ax_i^2)=x_i(2-ax_i). \quad (5.13)$$

用沒有除法装置的計算机时，有时可以用(5.13)計算倒数。又在求正規方陣 A 的逆矩陣 A^{-1} 时，有时也使用与(5.13)类似的公式

$$C_{i+1}=C_i(2-AC_i)\ ❸. \quad (5.14)$$

例5 計算 $1/\sqrt{a}$。令 $f(x)=1/x^2-a$，那么，对应的(5.9)是

$$x_{i+1}=x_i-(1/x_i^2-a)/(-2/x_i^3)=x_i+(x_i-ax_i^3)/2$$
$$=x_i(3-ax_i^2)/2. \quad (5.15)$$

这一公式的优点在于不包含除法。

作为高次代数方程式的解法，除了上述的方法外也許还应該提到 Graeffe 方法和 Bernoulli 方法等，但因为頁数的限制只好割爱。

---

❶ Barlow's Table of Squares, Cubes, etc 是著名的表。此外，Marchant 以及其他計算机制造公司推荐用1頁的表和1次除法(或乘法)算出5位的近似値，然后再用(5.12)加以改善而得到9位的近似値的方法，并且公布了为此所需要的表。

❷ 在本丛书森口著《穿孔卡計算机》第73頁例2中，取 $x_0=1$，不过，这样也沒有多大的差別。

❸ H. Hotelling: Proc. Berkerley Symp. Math. Stat. Prob. (1949), 275~293.

第2章　差分与插补

§6 差　分　表

函数 $f(x)$ 在一系列等距离(距离为 h)排列着的点 $x_i(i=\cdots, -2, -1, 0, 1, 2, \cdots)$ 的值已給定时,把

$$\Delta f(x)=f(x+h)-f(x) \tag{6.1}$$

叫做 $f(x)$ 的差分(递差 difference)。与微分是极其微小的差相反,因为差分是有限的差,因此也叫做有限差(finite difference)。

例 1　$f(x)=x^2$, $h=1$ 的情形。

表 6.1 中的 Δ 列表示 $\Delta f(x)$,对于它再施行差分演算得到 $\Delta^2 f(x)$,更进一步得到 $\Delta^3 f(x)$,在这种情形,$\Delta^3 f(x)$ 常等于 0 的事实并不是偶然的。即一般地有

$$\Delta f(x)=(x+1)^2-x^2=2x+1,$$
$$\Delta^2 f(x)=\{2(x+1)+1\}-(2x+1)=2,$$
$$\Delta^3 f(x)=2-2=0.$$

表 6.1　x^2 的差分

x	x^2	Δ	Δ^2	Δ^3
0	0			
		1		
1	1		2	
		3		0
2	4		2	
		5		0
3	9		2	
		7		0
4	16		2	
		9		0
5	25		2	
		11		0
6	36		2	
		13		
7	49			

給定 $f(x)$ 的表时，作出列举它的逐次差分的差分表(遞差表)，有时可以发現表的錯誤。

例 2　假設給定了如表 6.2 第1列，第2列所示的立方表。粗略地一看可能不会发現它的錯誤，但如作出差分表，就可以看出在 $x=6$ 的地方有可疑之处。应該是 $6^3=216$，而表內却是 $6^3=217$。

表 6.2　依据差分表的檢查

x	x^3	Δ	Δ^2	Δ^3	Δ^4
0	0				
		1			
1	1		6		
		7		6	
2	8		12		0
		19		6	
3	27		18		0
		37		6	
4	64		24		1
		61		7	
5	125		31		−4
		92		3	
6	217		34		6
		126		9	
7	343		43		−4
		169		5	
8	512		48		1
		217		6	
9	729		54		0
		271		6	
10	1000		60		
		331			
11	1331				

［注］　一般說来，$\Delta f(x_i)=f(x_{i+1})-f(x_i)$，$\Delta^2 f(x_i)=f(x_{i+2})-2f(x_{i+1})+f(x_i)$，$\Delta^3 f(x_i)=f(x_{i+3})-3f(x_{i+2})+3f(x_{i+1})-f(x_i)$，$\Delta^4 f(x_i)=f(x_{i+4})-4f(x_{i-3})+6f(x_{i+2})-4f(x_{i+1})+f(x_i)$，因此，如果 $f(x_i)$ 发生誤差 ε，那么，$\Delta^4 f(x_i)$，$\Delta^4 f(x_{i-1})$，$\Delta^4 f(x_{i-2})$，$\Delta^4 f(x_{i-3})$，$\Delta^4 f(x_{i-4})$ 就会分別发生誤差 $+\varepsilon$，-4ε，$+6\varepsilon$，-4ε，ε。在形如表 6.2 的差分表中，具有誤差 $+6\varepsilon$ 的 $\Delta^4 f(x_{i-2})$ 与 $f(x_i)$ 写在同一水平綫上，因此容易发現产生錯誤的地方。

卽使沒有錯誤，如果在函数值中有舍入誤差，同样也会逐漸扩大，另一方面"眞正的遞差"逐漸減小，因此不久就会出現不規則的

状态。如果把 $f(x_i)$ 中所含的舍入誤差看作是相互独立的具有相等方差 σ^2 的随机变量，而且眞正的 $\Delta^4 f(x)$ 小到足以无視的程度，那么，实际的 $\Delta^4 f(x_i)$ 的序列具有如下的自相关函数：

$$V(\Delta^4 f(x_i)) = \left\{\binom{4}{0}^2 + \binom{4}{1}^2 + \binom{4}{2}^2 + \binom{4}{3}^2 + \binom{4}{4}^2\right\}\sigma^2$$

$$= \binom{8}{4}\sigma^2 = 70\sigma^2,$$

$$V(\Delta^4 f(x_i),\ \Delta^4 f(x_{i+1}))$$

$$= -\left\{\binom{4}{0}\binom{4}{1} + \binom{4}{1}\binom{4}{2} + \binom{4}{2}\binom{4}{3} + \binom{4}{3}\binom{4}{4}\right\}\sigma^2$$

$$= -\binom{8}{3}\sigma^2 = -56\sigma^2,$$

$$V(\Delta^4 f(x_i),\ \Delta^4 f(x_{i+2}))$$

$$= \left\{\binom{4}{0}\binom{4}{2} + \binom{4}{1}\binom{4}{3} + \binom{4}{2}\binom{4}{4}\right\}\sigma^2 = \binom{8}{2}\sigma^2 = 28\sigma^2,$$

$$V(\Delta^4 f(x_i),\ \Delta^4 f(x_{i+3}))$$

$$= -\left\{\binom{4}{0}\binom{4}{3} + \binom{4}{1}\binom{4}{4}\right\}\sigma^2 = -\binom{8}{1}\sigma^2 = -8\sigma^2,$$

$$V(\Delta^4 f(x_i),\ \Delta^4 f(x_{i+4})) = \binom{4}{0}\binom{4}{4}\sigma^2 = \binom{8}{0}\sigma^2 = \sigma^2,$$

$$V(\Delta^4 f(x_i),\ \Delta^4 f(x_{i+k})) = 0 \qquad (k \geqslant 5).$$

由此推出，自相关系数是 $\rho_1 = -0.8$，$\rho_2 = +0.4$，$\rho_3 = -0.114$，$\rho_4 = +0.014$，$\rho_k = 0\,(k \geqslant 5)$。因为 ρ_1 是相当大的負值，通常在这种情形 $\Delta^4 f(x_i)$ 有正負相間的符号：$+ - + - \cdots$。还有，如果原来的函数值例如在小数5位作了舍入，那么，$\sigma^2 = 10^{-10}/12$，因此 $\Delta^4 f(x_i)$ 的标准离差是

$$D(\Delta^4 f(x_i)) = \sqrt{\frac{70}{12}} \times 10^{-5} = 2.4 \times 10^{-5}.$$

也就是說，以小数第5位为单位等于2.4。它的2倍是4.8，3倍是7.2，因此，可以认为 $\Delta^4 f(x_i)$ 通常在 ± 4 以內，偶然也有超过

±5, ±6 的时候，但极少超过 ±7 的时候。

例 3 $f(x)=\sin x$, $h=0.01$ 的情形。差分表如表 6.3 所示。可以认为 Δ^3, Δ^4 列只是含入誤差的影响。

表 6.3 $\sin x$ 的差分表

x	$f(x)$	Δ	Δ^2	Δ^3	Δ^4
0.90	0.78333				
		617			
0.91	0.78950		−7		
		610		−1	
0.92	0.79560		−8		1
		602		0	
0.93	0.80162		−8		0
		594		0	
0.94	0.80756		−8		−1
		586		−1	
0.95	0.81342		−9		3
		577		2	
0.96	0.81919		−7		−4
		570		−2	
0.97	0.82489		−9		3
		561		1	
0.98	0.83050		−8		−2
		553		−1	
0.99	0.83603		−9		
		544			
1.00	0.84147				

［注］ 一般来說，依据在小数 s 位作了含入的函数表所作的差分表中，由含入誤差引起的 n 级差的标准离差是

$$D(\Delta^n f(x_i)) = \sqrt{\frac{1}{12}\binom{2n}{n}} \times 10^{-s}. \qquad (6.2)$$

表 6.4 所示是右边第一因子对各种 n 的値求得的結果。

表 6.4 由含入誤差引起的差分的标准离差

n	1	2	3	4	5	6	7	8	9	10	11	12
$\sqrt{\frac{1}{12}\binom{2n}{n}}$	0.41	0.71	1.3	2.4	4.6	8.8	17	33	63	124	243	475

以小数第 s 位为单位，第 n 差分如果大体上在表内所列值的 2 倍以內(即使偶然超过 2 倍，但不超过 3 倍)，而且显著地有符号正負相間的傾向，那

么可以认为几乎只有誤差的影响❶。

§7 应用差分的插补公式

如果引用輔助变量 $u=\dfrac{x-x_0}{h}$，那么 $\Delta u=1$，对应于 $x=\cdots x_{-2}$, x_{-1}, x_0, x_1, x_2, $\cdots$ 的值是 $u=\cdots$, -2, -1, 0, 1, 2, $\cdots$. 反之，x 由 $x=x_0+hu$ 給出。

这样，如果定义"阶乘函数"为❷

$$u^{(m)}=u(u-1)\cdots(u-m+1),\tag{7.1}$$

那么，对于任意的正整数 m，下列公式成立：

$$\begin{aligned}\Delta u^{(m)}&=(u+1)u(u-1)\cdots(u-m+2)\\&\quad-u(u-1)\cdots(u-m+2)(u-m+1)\\&=mu(u-1)\cdots(u-m+2)=mu^{(m-1)}.\end{aligned}\tag{7.2}$$

其中令 $u^{(0)}\equiv1$。

如果 $f(x)$ 可以展开为

$$f(x)=f(x_0+hu)=a_0+a_1u^{(1)}+a_2u^{(2)}+a_3u^{(3)}+a_4u^{(4)}+\cdots,\tag{7.3}$$

那么，

$$\left.\begin{aligned}\Delta f(x)&=a_1+2a_2u^{(1)}+3a_3u^{(2)}+4a_4u^{(3)}+\cdots,\\\Delta^2 f(x)&=2a_2+6a_3u^{(1)}+12a_4u^{(2)}+\cdots,\\\Delta^3 f(x)&=6a_3+24a_4u^{(1)}+\cdots,\\\Delta^4 f(x)&=24a_4+\cdots,\\&\cdots\cdots\cdots\cdots\cdots\cdots\end{aligned}\right\}\tag{7.4}$$

在此如果令 $u=0$，那么 $u^{(m)}=0\ (m\geq1)$，因此就得到

$$f(x_0)=a_0,\quad \Delta f(x_0)=a_1,\quad \Delta^2 f(x_0)=2a_2,\quad \Delta^3 f(x_0)=6a_3,\quad \Delta^4 f(x_0)=24a_4,\ \cdots\tag{7.5}$$

---

❶ 关于差分的概率分布有 Miller 的研究[Math. Tables and other Aids to Computation, 4 (1950)]。

❷ 可以讀作"u 的 m 阶乘"。

为簡单計把这些左边分别用 f_0, Δ_0, Δ_0^2, Δ_0^3, Δ_0^4, … 表示，而就 a_0, a_1, … 解出，就得到

$$a_0=f_0,\quad a_1=\Delta_0,\quad a_2=\frac{\Delta_0^2}{2},\quad a_3=\frac{\Delta_0^3}{6},\quad a_4=\frac{\Delta_0^4}{24},\cdots. \tag{7.6}$$

代入(7.3)就得到

$$\begin{aligned} f(x)&=f(x_0+hu)\\ &=f_0+\Delta_0u^{(1)}+\frac{\Delta_0^2}{2}u^{(2)}+\frac{\Delta_0^3}{6}u^{(3)}+\frac{\Delta_0^4}{24}u^{(4)}+\cdots, \end{aligned}\tag{7.7}$$

这就是所謂 **Newton 前进插补公式**。

式中 Δ_0^m 的分母可以写作 $m!$，又因为 $u^{(m)}/m!$ 可以写作 $\binom{u}{m}$，因此(7.7)也可以写成

$$\begin{aligned} f(x)&=f(x_0+hu)\\ &=f_0+\Delta_0\binom{u}{1}+\Delta_0^2\binom{u}{2}+\Delta_0^3\binom{u}{3}+\Delta_0^4\binom{u}{4}+\cdots. \end{aligned}\tag{7.8}$$

例 1　应用表 6.3 求 sin 0.923.

令 $x_0=0.92$, $h=0.01$, $u=0.3$ 而应用(7.7)就得到

$$\begin{aligned} \sin 0.923=0.79560&\\ +\Big[&602\times0.3+(-8)\frac{0.3\times(-0.7)}{2}+0\frac{0.3\times(-0.7)\times(-1.7)}{6}\\ &+(-1)\frac{0.3\times(-0.7)\times(-1.7)\times(-2.7)+\cdots}{24}\Big]\times10^{-5}\\ =0.79560&+(180.6+0.84+0+0.04+\cdots)\times10^{-5}\\ \approx0.79741. \end{aligned}$$

［注］　如前所述，Δ_0^3, Δ_0^4 可能只是舍入誤差的影响，因此含有这些项几乎沒有意义。根据更詳細的表，sin 0.923=0.79741 5498；因此，上面求得的值只不过在最后一位有不超过 1 的誤差而已。

例 2　应用公式(7.8)和表 7.1 計算 $\phi(0.634)$.

如表 7.2 所示，得到 $\phi(0.634)=0.32630\ 70293$. 根据更詳細的表 $\phi(0.634)=0.32630\ 70292\ 88456$，由此看出，在第 10 位有不超过 1 的誤差。

和(7.8)的推导类似，可以推出如下的两个公式：

$$f(x)=f(x_0+hu)$$
$$=f_0+\Delta_0\binom{u}{1}+\Delta^2_{-1}\binom{u}{2}+\Delta^3_{-1}\binom{u+1}{3}+\Delta^4_{-2}\binom{u+1}{4}+\cdots,\quad(7.9)$$
$$f(x)=f(x_0+hu)$$
$$=f_0+\Delta_{-1}\binom{u}{1}+\Delta^2_{-1}\binom{u+1}{2}+\Delta^3_{-2}\binom{u+1}{3}+\Delta^4_{-2}\binom{u+2}{4}+\cdots.\quad(7.10)$$

它們分别叫做 Gauss 前进插补公式和后退插补公式。

表 7.1 $\phi(x)=(2\pi)^{-\frac{1}{2}}\exp(-x^2/2)$ 的差分表

x	$\phi(x)$	Δ	Δ^2	Δ^3	Δ^4
0.60	0.33322 46029				
		−200 99227			
0.61	0.33121 46802		−2 07967		
		−203 07194		5323	
0.62	0.32918 39608		−2 02644		30
		−205 09838		5353	
0.63	0.32713 29770		−1 97291		23
		−207 07129		5376	
0.64	0.32506 22641		−1 91915		24
		−208 99044		5400	
0.65	0.32297 23597		−1 86515		21
		−210 85559		5421	
0.66	0.32086 38038		−1 81094		18
		−212 66653		5439	
0.67	0.31873 71385		−1 75655		17
		−214 42308		5456	
0.68	0.31659 29077		−1 70199		14
		−216 12507		5470	
0.69	0.31443 16570		−1 64729		
		−217 77236			
0.70	0.31225 39334				

表 7.2 $\phi(0.634)$ (Newton 式)

m	差分	插补系数	乘积
0	0.32713 29770	1	0.32713 29770
1	−207 07129	0.4	−82 82851 6
2	−1 91915	−0.12	23029 8
3	5400	0.064	345 6
4	21	−0.0416	−8736
总计			0.32630 70292 9264

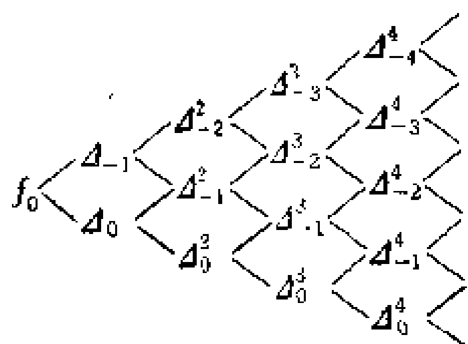


图7.1 菱形图形

[注] 公式的要点是，一般地，如果Δ^m的足号为$-s_m$，那么Δ^{m+1}后面的二項系数是$\binom{u+s_m}{m+1}$，以及在图7.1中沿着由斜綫組成的鏈向右前进(或者也可以說，对于$m=0, 1, 2, \cdots$，s_{m+1}与s_m或s_m+1相等)。按照这法則(Sheppard 法則)所作的每一个展开式都可以和(7.8)类似地推得。在图7.1中，写着按照上述法則确定的二項系数的图形叫做菱形图形(lozenge diagram)。

在(7.9)，(7.10)中，使用通过将要插补的点的水平綫附近的差分是它的特征。如果取二者的平均数，就得到关于$u=0$对称的如下的 **Stirling 插补公式**：

$$f(x)=f(x_0+hu)=f_0+\frac{\Delta_{-1}+\Delta_0}{2}u+\Delta^2_{-1}\frac{u^2}{2}$$
$$+\frac{\Delta^3_{-2}+\Delta^3_{-1}}{2}\frac{u(u^2-1)}{6}+\Delta^4_{-2}\frac{u^2(u^2-1)}{24}+\cdots. \quad (7.11)$$

如果把(7.10)中各差分的足号都加1，并用u代替$u-1$，把这样得到的公式与(7.9)平均，就得到关于$u=\frac{1}{2}$对称的插补公式。将这公式用新变量$v=u-\frac{1}{2}$表出就得

$$f(x)=\frac{f_0+f_1}{2}+\Delta_0 v+\frac{\Delta^2_{-1}+\Delta^2_0}{2}\frac{1}{2}\left(v^2-\frac{1}{4}\right)+\Delta^3_{-1}\frac{v}{6}\left(v^2-\frac{1}{4}\right)$$
$$+\frac{\Delta^4_{-2}+\Delta^4_{-1}}{2}\frac{1}{24}\left(v^2-\frac{1}{4}\right)\left(v^2-\frac{9}{4}\right)+\cdots. \quad (7.12)$$

这公式叫做 **Bessel 插补公式**。

如果有(7.9)～(7.12)的对应的插补系数表，計算的繁簡都差不多，精确度也大体上相同。

其次，在(7.9)中代入$\Delta_0=f_1-f_0$，$\Delta^3_{-1}=\Delta^2_0-\Delta^2_{-1}$，$\Delta^5_{-2}=\Delta^4_{-1}-\Delta^4_{-2}$，…以消去奇数級的差分，就得到

$$f(x)=f(x_0+hu)=(1-u)f_0+uf_1-\frac{u(u-1)(u-2)}{6}\Delta^2_{-1}$$
$$+\frac{(u+1)u(u-1)}{6}\Delta^2_0-\frac{(u+1)u(u-1)(u-2)(u-3)}{120}\Delta^4_{-2}$$
$$+\frac{(u+2)(u+1)u(u-1)(u-2)}{120}\Delta^4_{-1}+\cdots.$$

在这公式中令 $u'=1-u$, $E_2=u(1-u^2)/6$, $E'_2=u'(1-u'^2)/6$, $E_4=u(1-u^2)(4-u^2)/120$, $E'_4=u'(1-u'^2)(4-u'^2)/120,\cdots$就得到

$$f(x)=(u'f_0+uf_1)-(E'_2\Delta^2_{-1}+E_2\Delta^2_0)$$
$$+(E'_4\Delta^4_{-2}+E_4\Delta^4_{-1})+\cdots. \qquad (7.13)$$

这公式叫做 **Everett 插补公式。** 其特征是使用着直綫 $u=0$ 和 $u=1$ 上的(偶数級的)差分,在标明着这些差分的表中,使用这公式最为便利。在这样的表中多半附有 E_2, E'_2, E_4, E'_4 的表。此外,如果使用中心差分(central difference)的記号

$$\delta f(x)=f\left(x+\frac{h}{2}\right)-f\left(x-\frac{h}{2}\right), \qquad (7.14)$$

而且把 $\delta^m f(x_i)$ 簡写作 δ^m_i, 那么(7.13)就成为

$$f(x)=(u'f_0+uf_1)-(E'_2\delta^2_0+E_2\delta^2_1)$$
$$+(E'_4\delta^4_0+E_4\delta^4_1)+\cdots, \qquad (7.15)$$

这样,式子的形状得到进一步的簡化。

例 3 应用 Everett 插补公式(7.15)和表 7.3 求 $\phi(0.634)$.

表7.3 $\phi(x)$ 的 表

x	$\phi(x)$	δ^2	δ^4
0.63	0.32713 29770	−1 97291	23
0.64	0.32506 22641	−1 91915	24

$u'f_0+uf_1=$	0.32630 46918 4
$-(E'_2\delta^2_0+E_2\delta^2_1)=$	23373 864
$E'_4\delta^4_0+E_4\delta^4_1=$	52595
总 計	=0.32630 70292 78995

因为 $u=0.4$, $u'=0.6$, $E_2=0.056$, $E_2'=0.064$, $E_4=0.010752$, $E_4'=0.011648$, 如表 7.3 那样进行計算,就得到 $\phi(0.634)=0.3263070293$.

［注］ 为了使用 L. J. Comrie 所提出的 "throw back" 方法，有記載着"修正差分" $\delta^{2*}=\delta^2-c\delta^4$ 的表 (例如 Nat. Bur. Stds., Tables of Bessel Functions of Fractional Order, vol. 1. 1948)．应用 δ^{2*} 計算

$$f(x)=(u'f_0+uf_1)-(E_2'\delta_0^{2*}+E_2\delta_1^{2*}), \tag{7.16}$$

可以得到和在(7.15)中取到 δ^4 的項时同样的精确度。c 可以取作 0.184. 如果在上面的例中使用这样的 c 的值,在表中代替 δ^2 应該加上 $\delta_0^{2*}=-197295$, $\delta_1^{2*}=-191919$, 由此可以求得 $-(E_2'\delta_0^{2*}+E_2\delta_1^{2*})=0.00000\,23374\,344$, 把它和 $u'f_0+uf_1$ 相加作为答案。

由于这一置換引起的誤差是 $E=(cE_2'-E_4')\delta_0^4+(cE_2-E_4)\delta_1^4=\{c(E_2'+E_2)-(E_4'+E_4)\}\delta_0^4+(cE_2-E_4)\delta_{\frac{1}{2}}^5/2$, 但由于取 $c=0.184$ 时 $c(E_2'+E_2)-(E_4'+E_4)=cu(1-u)/2-u(1-u^2)(2-u)/24$ 的絕对値在 $0\leqslant u\leqslant 1$ 的范圍內不超过 0.00045, 而且 $|cE_2-E_4|$ 不超过 0.0008, 因此,如果 $|\delta^4|<1000$, $|\delta^5|<60$, 那么 $|E|<0.5$ (参看 §13 例 1)．

§8 Lagrange 插补公式

在前节中討論的插补公式都是应用差分的插补公式。与此相反，将函数値自身乘以适当的系数而后相加从而得到插补値的是 Lagrange 插补公式。例如,在"5 点法"中利用 $f_i=f(x_i)$ $(i=-2, -1, 0, 1, 2)$, 而設

$$f(x)=f(x_0+hu)=A_{-2}f_{-2}+A_{-1}f_{-1}+A_0f_0+A_1f_1+A_2f_2, \tag{8.1}$$

其中的系数——Lagrange 插补系数——A_i 与函数 $f(x)$ 无关而只由 u 确定,如果备有由 u 查 A_i 的表[1]，使用起来是便利的。

在 Gauss 前进插补公式 (7.9) 中取至含 Δ^4 的項，并把插补系数表示成

---

[1] 例如, Marchant Methods. MM-228, 1942; NBS, Tables of Lagrangian Interpolation Coefficients, Col. Univ. Press (1944)．石田保士,补間系数表,培风館(1953)．此外, u 的步长为 1/60 的表有 NBS. AMS. No. 35. 1954.

$$G_0=1,\ G_1=u,\ G_2=u(u-1)/2,\ G_3=(u+1)u(u-1)/6,$$

$$G_4=(u+1)u(u-1)(u-2)/24, \tag{8.2}$$

并在各差分的地方代入

$$\Delta_0=f_1-f_0,\ \Delta^2_{-1}=f_1-2f_0+f_{-1},\ \Delta^3_{-1}=f_2-3f_1+3f_0-f_{-1},$$

$$\Delta^4_{-2}=f_2-4f_1+6f_0-4f_{-1}+f_{-2}, \tag{8.3}$$

然后整理成(8.1)的形状就得到

$$A_{-2}=G_4,\ A_{-1}=G_2-G_3-4G_4,\ A_0=1-G_1-2G_2+3G_3+6G_4,$$

$$A_1=G_1+G_2-3G_3-4G_4,\ A_2=G_3+G_4. \tag{8.4}$$

将(8.2)代入(8.4)就得到

$$\left.\begin{aligned} A_{-2}&=(u+1)u(u-1)(u-2)/24,\\ A_{-1}&=(u+2)u(u-1)(u-2)/(-6),\\ A_0&=(u+2)(u+1)(u-1)(u-2)/4,\\ A_1&=(u+2)(u+1)u(u-2)/(-6),\\ A_2&=(u+2)(u+1)u(u-1)/24. \end{aligned}\right\} \tag{8.5}$$

显然可以看出，这里引入的 A_i 当 $u=j$ 时成为 $\delta_{ij}=1(i=j)$，$=0(i\neq j)$。此外，也容易明白具有上述性质的 4 次式是唯一的。

例 1 利用表 7.1 的函数值，由 Lagrange 插补公式(8.1)求 $\phi(0.634)$。

令 $x_0=0.63$，$h=0.01$，$u=0.4$。对于 $u=0.4$，$A_{-2}=0.0224$，$A_{-1}=-0.1536$，$A_0=0.8064$，$A_1=0.3584$，$A_2=-0.0336$，因此，如表 8.1 所示的那样得到 $\phi(0.634)=0.32630\ 70293$。

［注］ 用台式計算机作形如 $\sum A_i f_i$ 的計算时，可以不必作中間記录而用一連串的操作来完成它。在能够同时作出乘数 A_i 的和的計算机械中，查明 $\sum A_i=1$ 的事实可以作为驗算。

如同在上面看到的那样，在理論上，利用系数(8.5)的 Lagrange 公式(8.1)和取至含 Δ^4 的

表 8.1 依据 Lagrange 公式計算 $\phi(0.634)$

x	$\phi(x)$	插补系数
0.61	0.33121 46802	+0.0224
0.62	0.32918 39608	−0.1536
0.63	0.32713 29770	+0.8064
0.64	0.32506 22641	+0.3584
0.65	0.32297 23597	−0.0336
$\phi(0.634)=0.32630\ 70292\ 77920$		

項为止的 Gauss 前进公式(7.9)完全是等价的，因此，如果在計算过程中沒有作舍入，就应該得到完全相同的結果。但是，如果插补系数作了舍入，那就是另外一回事了。例如，对于 $u=0.41$ 計算(8.5)，到小数第 7 位止作舍入，就得到表 8.2 的第 2 列的值，用这些值計算 $\phi(0.6341)$，就得到值 0.32628 63731 48…。另一方面，把 Gauss 插补系数(8.2)各到小数第 7 位舍入(不过，在本例中，有尾数的只有 G_4 一項)，代入(8.4)求 A_i，就得到表 8.2 第 3 列那样的值，由此求得的 $\phi(0.6341)$ 的值是 0.32628 63404 36…。根据精密的表 $\phi(0.6341)=0.32628\ 63404\ 47\cdots$，因此，后面的計算法給出滿意的結果，而由前面的計算法得到的結果包含着 327×10^{-10} 左右的誤差。这一誤差主要是由于 $\sum A_i$ 不等于 1 而在第 7 位有誤差 +1 所引起的。在第 42 頁的脚注中提到的旧式表中大多都有这种缺点。现在已經有了沒有这种缺点的、由后面的方法求得的、步长为 0.01 的 A_i 的表❶。步长 0.001 的表目前正在准备付印。

表 8.2　A_i 的舍入方法的比較($u=0.41$)

i	分別作了舍入的值	把舍入了的 G_m 代入(8.4)所得的值
−2	0.0225 965	0.0225 965
−1	−0.1544 894	−0.1544 895
0	0.7969 394	+0.7969 395
1	0.3692 036	+0.3692 035
2	−0.0342 500	−0.0342 500
总計	1.0000 001	1.0000 000

Lagrange 插补公式的优点是，在表中使用函数值自身，因而不必要計算差分表。但是，另一方面，究竟用几点法比較恰当的問題不能由計算本身决定，这是它的缺点，因此，例如使用了 5 点法时就要計算 $\Delta^5_{-2}=f_3-5f_2+10f_1-10f_0+5f_{-1}-f_{-2}$，看它的絕对值是否超过 40 (因为 G_5 不超过 0.012，因此，如果 $|\Delta^5|<40$，那么 $|G_5\Delta^5|<0.5$)。如果 $\Delta^4_{-2}=f_2-4f_1+6f_0-4f_{-1}+f_{-2}$ 的絕对值不超过 20，那么用 4 点法也可以。如果 $\Delta^3_{-1}=f_2-3f_1+3f_0-f_1$ 的絕对值不超过 8，那么用 3 点法也可以。如果 $\Delta^2_{-1}=f_1-2f_0+f_1$ 的絕对值不超过 4，那么用 2 点法(綫性插补，即普通的比例部分方法)

❶ Sigeiti Moriguti (森口繁一): Rep. Stat. Appl. Res., JUSE, 4(1957), 37～42.

也可以。

在給定函数值的点 $x_i(i=0, 1, 2, \cdots)$ 不是等距离的情形，也容易写出 Lagrange 插补公式。例如，在 4 点法的情形，

$$f(x)=\frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}f_0+\frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}f_1$$
$$+\frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}f_2+\frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}f_3.$$

(它是满足 $f(x_i)=f_i(i=0, 1, 2, 3)$ 的唯一 3 次式。)

在一般情形，設 $\Pi(x)=(x-x_0)(x-x_1)\cdots(x-x_n)$，那么，$(n+1)$点法的 Lagrange 插补公式可以写成

$$f(x)=\sum_{i=0}^{n}\frac{\Pi(x)}{(x-x_i)\Pi'(x_i)}f_i. \tag{8.6}$$

在作实际計算时，可以如表 8.3 那样，計算

$$\Pi(x)=(x-x_0)(x-x_1)(x-x_2)(x-x_3),$$
$$D_0=(x-x_0)(x_0-x_1)(x_0-x_2)(x_0-x_3),$$
$$D_1=(x_1-x_0)(x-x_1)(x_1-x_2)(x_1-x_3),$$
$$D_2=(x_2-x_0)(x_2-x_1)(x-x_2)(x_2-x_3),$$
$$D_3=(x_3-x_0)(x_3-x_1)(x_3-x_2)(x-x_3),$$

然后設 $S=f_0/D_0+f_1/D_1+f_2/D_2+f_3/D_3$，因而 $f(x)=\Pi(x)\cdot S$ 而求插补值是便利的。

表 8.3　(8.6) 的計算

$x-x_0$	x_0-x_1	x_0-x_2	x_0-x_3	D_0	f_0	f_0/D_0
x_1-x_0	$x-x_1$	x_1-x_2	x_1-x_3	D_1	f_1	f_1/D_1
x_2-x_0	x_2-x_1	$x-x_2$	x_2-x_3	D_2	f_2	f_2/D_2
x_3-x_0	x_3-x_1	x_3-x_2	$x-x_3$	D_3	f_3	f_3/D_3
$\Pi(x)=\cdots\cdots$						S
$f(x)=\Pi(x)\cdot S=\cdots\cdots$						

表 8.4

x	$f(x)$
0.74765 715	0.642
0.74921 731	0.643
0.75077 981	0.644
0.75234 466	0.645

例 2　由表 8.4 求 $f(0.75000\ 000)$.

如表 8.5 那样进行計算，得到插补值是 0.64350 11090.

表 8.5　非等距离 Lagrange 插补公式計算的例子

0.00234 285	−0.00156 016	−0.00312 266	−0.00468 751
0.00156 016	0.00078 269	−0.00156 250	−0.00312 735
0.00312 266	0.00156 250	−0.00077 981	−0.00156 485
0.00468 751	0.00312 735	0.00156 485	−0.00234 466

$\Pi(x) = 3.35276\ 36830 \times 10^{-12}$

D_i	f_i	f_i/D_i
$-53.50829\ 520 \times 10^{-12}$	0.642	$-0.01199\ 92609 \times 10^{12}$
$+5.95699\ 1765 \times 10^{-12}$	0.643	$+0.10775\ 94917 \times 10^{12}$
$+5.95396\ 4495 \times 10^{-12}$	0.644	$+0.10816\ 32247 \times 10^{12}$
$-53.78625\ 224 \times 10^{-12}$	0.645	$-0.01199\ 19119 \times 10^{12}$
$f(x) = 0.64350\ 11090$		$S = 0.19193\ 15436 \times 10^{12}$

〔注〕 表 8.4 实际上是反正切 ($\mathrm{arctg}\, x$) 的表。因为它是把正切表中自变量的值記入 $f(x)$ 栏、把正切的值記入 x 栏而得到的。因此，目前求得的值就是 arctg(0.75) 的近似値。依据到小数 12 位的反正切表，arctg(0.75) $=0.64350\ 11087\ 93$，因此，目前求得的值到小数第 9 位大体上是精确的。

非等距离的 Lagrange 插补公式常常象上例那样用于反插补 (inverse interpolation) 問題。

Aitken 創造了实行非等距离插补的逐次計算方法[1]，即当給定了 $f_i = f(x_i)$ $(i=0, 1, 2, \cdots)$ 时，首先求

$$I_{0i}(x) = \frac{1}{x_i - x_0}\begin{vmatrix} f_0 & x_0 - x \\ f_i & x_i - x \end{vmatrix} \quad (i=1, 2, \cdots), \tag{8.7}$$

其次求

$$I_{01i}(x) = \frac{1}{x_i - x_1}\begin{vmatrix} I_{01}(x) & x_1 - x \\ I_{0i}(x) & x_i - x \end{vmatrix} \quad (i=2, 3, \cdots), \tag{8.8}$$

再其次求

$$I_{012i}(x) = \frac{1}{x_i - x_2}\begin{vmatrix} I_{012}(x) & x_2 - x \\ I_{01i}(x) & x_i - x \end{vmatrix} \quad (i=3, 4, \cdots) \tag{8.9}$$

等等。

[1] A. C. Aitken: Proc. Edin. Math. Soc. (2), 3 (1932), 56～84.

显然，$I_{0i}(x)$ 是当 $x=x_0$ 时等于 f_0，而当 $x=x_i$ 时等于 f_i 的一次式。而且，由此容易証明，由 (8.8) 定义的 $I_{01i}(x)$ 是满足 $I_{01i}(x_0)=f_0$，$I_{01i}(x_1)=f_1$，$I_{01i}(x_i)=f_i$ 的 2 次式；由此并推出，(8.9) 是在 $x=x_0$，x_1，x_2，x_i 和 $f(x)$ 相等的 4 次式。这样就可以逐次計算出在实质上和 (8.6) 相同的插补式的值。这时，由于 x_0，x_1，x_2，… 的排列方法是任意的，因此可以由距离 x 最近的点开始順次取較远的点。至于計算的記录，写成表 8.6 的形状是便利的。

表 8.6 Aitken 的逐次計算法

x_0	f_0				x_0-x
x_1	f_1	$I_{01}(x)$			x_1-x
x_2	f_2	$I_{02}(x)$	$I_{012}(x)$		x_2-x
x_3	f_3	$I_{03}(x)$	$I_{013}(x)$	$I_{0123}(x)$	x_3-x

例 3 用 Aitken 逐次計算法計算例 2 中的問題。

如表 8.7 那样进行計算，可以看到，到 I_{01i} 即 3 点法时已經收斂，因此，沒有继續往下进行計算的必要。而且，这里得到的值大体上到小数第 9 位是正确的。

表 8.7 arctg(0.75) 的計算 (Aitken 方法)

i	x_i	f_i	I_{0i}	I_{01i}	x_i-x
0	0.75077 981	0.644			0.00077 981
1	0.74921 731	0.643	0.64350 09216		−0.00078 269
2	0.75234 466	0.645	0.64350 16711	0.64350 11092	0.00234 466
3	0.74765 715	0.642	0.64350 05476	0.64350 11092	−0.00234 285

[注 1] 这个方法的中間計算，几乎是相同程序的重复，而且記录下来的数值都有它的意义，因此可以由此防止錯誤的产生或判断收斂的程度等，这是 Aitken 方法的优点。

[注 2] Neville 提出了与上述方法略为不同的先求 $I_{01}(x)$，$I_{12}(x)$，$I_{23}(x)$，…，再求 $I_{012}(x)$，$I_{123}(x)$，… 的方法❶。还有，Tweedie 提出了在 x_i 按下标的次序排列的条件下，顺次計算 $I_{0i}(x)$ $(i=1, 2, \cdots)$，$I_{j1}(x)$ $(j=-1, -2, \cdots)$，$I_{01i}(x)$ $(i=2, 3, \cdots)$，$I_{j01}(x)$ $(j=-1, -2, \cdots)$，… 的方法❷。

---

❶ E. H. Neville: Journ. Indian Math., Soc., 20 (1934), 87~120.

❷ M. C. K. Tweedie: Math. Tables and other Aids to Comp., (8) (1954), 13~16.

第3章　数值积分·数值微分

§9 Newton-Cotes 数值积分公式

数值积分公式中最常用的是下列的 Simpson 法则:

$$\int_{x_0}^{x_2} f(x)\,dx=\frac{h}{3}(f_0+4f_1+f_2). \tag{9.1}$$

其中 $x_i=x_0+ih$ $(i=0,\ 1,\ 2)$，$f(x_i)=f_i$.

公式的推导　满足 $f(x_i)=f_i(i=0,\ 1,\ 2)$ 的二次式是

$$f(x)=f(x_0+hu)=\frac{(u-1)(u-2)}{2}f_0+\frac{u(u-2)}{-1}f_1+\frac{u(u-1)}{2}f_2.$$

将这 $f(x)$ 由 $x=x_0$ 到 $x=x_2$ (也就是由 $u=0$ 到 $u=2$) 积分，就得到

$$\int_{x_0}^{x_2} f(x)\,dx=h\Big[f_0\int_0^2\frac{(u-1)(u-2)}{2}\,du+f_1\int_0^2\frac{u(u-2)}{-1}\,du$$
$$+f_2\int_0^2\frac{u(u-1)}{2}\,du\Big]=\frac{h}{3}(f_0+4f_1+f_2).$$

这公式是可以用类似方法推得的下列 Newton-Cotes 公式在 $n=2$ 时的情形:

$$\int_{x_0}^{x_n} f(x)\,dx=Kh(A_0f_0+A_1f_1+\cdots+A_nf_n). \tag{9.2}$$

其中 $f_i=f(x_i)=f(x_0+ih)$ $(i=0,\ 1,\ \cdots,\ n)$. 系数 A_i $(i=0,\ 1,\ \cdots,\ n)$ 和 K 如表 9.1 所示。

$n=1$ 时的 Newton-Cotes 公式叫做梯形法则。$n=3$ 时的情形也叫做 Simpson 3/8 法则。又将 $n=6$ 时的 Newton-Cotes 公式稍加变形，就可以导出 Weddle 法则

$$\int_{x_0}^{x_6} f(x)\,dx=\frac{6h}{20}(f_0+5f_1+f_2+6f_3+f_4+5f_5+f_6). \tag{9.3}$$

这虽然是一个次数較高的公式，但却具有系数简单的优点。

(9.3)的推导　在 $n=6$ 时的 $\sum A_if_i$ 上加上 $\Delta^6=f_0-6f_1+15f_2-20f_3$

表 9.1　Newton-Cotes 公式的系数

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=8$	$n=10$	$n=12$
A_i	1	1	1	7	19	41	989	+16067	+13 64651
	1	4	3	32	75	216	5888	+1 06300	+99 03168
		1	3	12	50	27	−928	−48525	−75 87864
			1	32	50	272	10496	+2 72400	+357 25120
				7	75	27	−4540	−2 60550	−514 91295
					19	216	10496	+4 27368	+875 16288
						41	以下对称	−2 60550	−877 97136
								以下对称	+875 16288
									以下对称
$K=$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{3}{8}$	$\frac{2}{45}$	$\frac{5}{288}$	$\frac{1}{140}$	$\frac{4}{14175}$	$\frac{5}{2\,99376}$	$\frac{1}{52\,65250}$

在 Kopal 书(見本书末的文献[4])的p.p.536～538 有对于 $n=1$ (1) 20 的 A_i 和 $D=n/2K$ 的表❶ (这里的 K 可以由它的 D 按关系 $K=\frac{n}{2D}$ 求出)

$+15f_4-6f_5+f_6$ 就得到 $42(f_0+5f_1+f_2+6f_3+f_4+5f_5+f_6)$．也就是說，(9.3) 与 $n=6$ 时的 (9.2) 的差等于 $\Delta_0^6 h/140$，当 Δ_0^6 充分小时是可以略去不計的。

也可以将积分区間分成若干部分，而在每一部分中使用上述諸公式之一，然后将結果相加起来而求积分的近似値。这时候，各部分区間交界处的纵坐标使用两次，因此，将它的系数預先两倍起来而只計算一次是便利的。

例 1　按 Simpson 法則計算

$$\phi(x)=(2\pi)^{-\frac{1}{2}}\exp(-x^2/2)$$

由 0 到 1.2 的积分，而令 $h=0.1$。

計算可以如表 9.2 的記录那样进行。作出函数值与系数的乘积之和，用 3 除然后再乘以 $h=0.1$ (移动小数点) 就得到 0.33493 05322．按照詳細的表，积分的值

表 9.2　$\int_0^{1.2}\phi(x)dx$ 的計算

x	$\phi(x)$	系数
0	0.39894 22804	1
0.1	0.39695 25475	4
0.2	0.39104 26940	2
0.3	0.38138 78155	4
0.4	0.36827 01403	2
0.5	0.35206 53268	4
0.6	0.33322 46029	2
0.7	0.31225 39334	4
0.8	0.28969 15528	2
0.9	0.26608 52499	4
1.0	0.24197 07245	2
1.1	0.21785 21770	4
1.2	0.19418 60550	1
$h=0.1$		$K=1/3$

❶ 表的来源是 Johnson: Quart. Journ. Math., 46 (1915), 52.

是 0.38493 03297 78 ⋯，因此，上面得到的值的誤差是 2.024×10^{-7}。

为了弄明白用 Newton-Cotes 公式求得的积分值所含的誤差如何按次数 n 以及所使用的纵坐标的个数 N 等而变化，将例 1 中的积分用种种方法求出作比較，其結果如表 9.3 和图 9.1 所示。

表 9.3 Newton-Cotes 公式的誤差 E (例 1)

N \ n	1	2	3	4	5	6	Weddle	8	10	12
2	$-.029$									
3	$-.0^{2}71$	$+.0^{3}28$								
4	$-.0^{2}31$	—	$+.0^{3}12$							
5	$-.0^{2}18$	$+.0^{4}17$	—	$-.0^{6}57$						
6	$-.0^{2}11$	—	—	—	$-.0^{6}33$					
7	$-.0^{3}78$	$+.0^{5}33$	$-.0^{5}74$	—	—	$-.0^{7}19$	$-.0^{7}46$			
9	$-.0^{3}44$	$+.0^{5}10$	—	$-.0^{7}15$	—	—	—	$-.0^{9}5$		
11	$-.0^{3}28$	$+.0^{6}42$	—	—	$-.0^{8}82$	—	—	—	$-.0^{9}2$	
13	$-.0^{3}19$	$+.0^{6}20$	$+.0^{6}46$	$-.0^{8}14$	—	$.0^{10}$	$-.0^{8}7$	—	—	$+.0^{9}1$

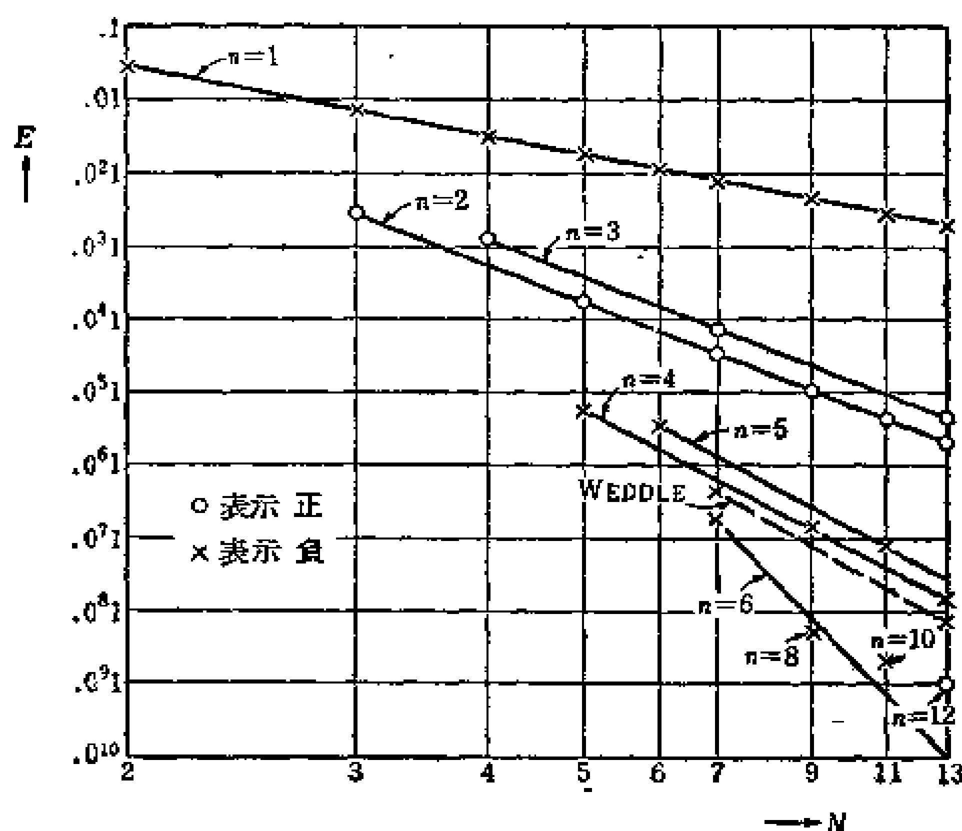


图 9.1 Newton-Cotes 公式的誤差(例 1)

全部計算都是按10位进行的。在使用 $N=13$ 个纵坐标和 $n=6$ 乃至 $n=12$ 的公式时，除了舍入誤差外沒有发生别的誤差。在其他的情形发生了某种程度的"截断誤差"(truncation error)。图9.1的纵軸是誤差絕对值的对数分度，而横軸是 $N-1$ 的对数分度❶。正的誤差和負的誤差分别用○或×的記号标出。对应于同一 n 的点分布在一条直綫上，由直綫的斜率可以看出，当 $n=1$ 时 $|E|\propto(N-1)^{-2}$，而当 $n=2$ 和 $n=3$ 时 $|E|\propto(N-1)^{-4}$。还可以看到，$n=3$ 时的 E 与 $n=2$ 时的 E 有相同的符号，而且前者常具有約为后者的2.3倍的大小。在这种情形，Weddle 法則比 $n=6$ 时的 Newton-Cotes 公式具有稍微大点的誤差。

作为在理論上处理 Newton-Cotes 公式的截断誤差方法的例子，考察 $n=2$，即 Simpson 法則的情形。精密地說，(9.1) 的右边是近似值而左边是精确值，因此，誤差是

$$E[f]=\frac{h}{3}\{f(x_0)+4f(x_1)+f(x_2)\}-\int_{x_0}^{x_2}f(x)dx. \quad (9.4)$$

这誤差由函数 $f(x)$ 确定，因此，可以說是 $f(x)$ 的泛函。因而可以形式地写成

$$E[f]=\int_{-\infty}^{\infty}g(x)f(x)dx, \quad (9.5)$$

其中

$$g(x)=\frac{h}{3}\{\delta(x-x_0)+4\delta(x-x_1)+\delta(x-x_2)\}$$
$$-\{1(x-x_0)-1(x-x_2)\}, \quad (9.6)$$

而 $\delta(x)$ 表示 Dirac 的 Delta 函数，$1(x)$ 表示单位阶梯函数 $[1(x)=1(x>0), =0(x<0)]$。現在，把 $g(x)$ 的逐次积分写成

$$g^{(-1)}(x)=\int_{-\infty}^{x}g(u)du,$$

$$g^{(-k-1)}(x)=\int_{-\infty}^{x}g^{(-k)}(u)du \quad (k=1, 2, \cdots) \quad (9.7)$$

(图9.2)，注意到当 $x<x_0$ 和 $x>x_2$ 时 $g(x)$，$g^{(-1)}(x)$，…，$g^{(-4)}(x)$

❶ 即图9.1中使用的是双对数坐标。——譯者注

都等于零，用分部积分法就可以把(9.5)变形如下：

$$E[f]=[g^{(-1)}(x)f(x)-g^{(-2)}(x)f'(x)+g^{(-3)}(x)f''(x)$$
$$-g^{(-4)}(x)f'''(x)]_{-\infty}^{\infty}+\int_{-\infty}^{\infty}g^{(-4)}(x)f^{(4)}(x)\,dx$$
$$=\int_{-\infty}^{\infty}g^{(-4)}(x)f^{(4)}(x)\,dx=\int_{x_0}^{x_2}g^{(-4)}(x)f^{(4)}(x)\,dx. \quad (9.8)$$

但是，由于 $g^{(-4)}(x)\geqslant 0\ (x_0<x<x_2)$，因此，根据中值定理❶

$$E[f]=f^{(4)}(\xi)\int_{x_0}^{x_2}g^{(-4)}(x)\,dx. \quad (9.9)$$

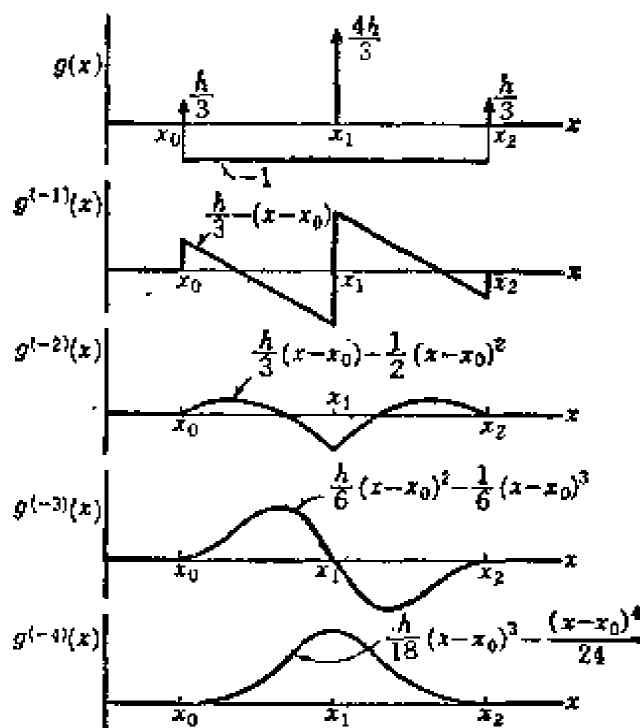


图 9.2　$g(x)$ 和它的逐次积分

其中 ξ 是积分区間 (x_0, x_2) 内的某一点。(9.9)右边定积分的值是 $h^5/90$，因此，最后得到

$$E[f]=\frac{h^5}{90}f^{(4)}(\xi). \quad (9.10)$$

用这种方法对于 n 的各种值求得(近似值-精确值)的誤差公式总括起来如表 9.4 所示（表中略去了导数中的自变量值 ξ）.

❶　例如，高木貞治：解析概論(增訂版)(岩波，1943) p. 113，定理 33.

表 9.4　Newton-Cotes 公式的誤差(一个区間)

n	1	2	3	4	5	6	8	10
誤差	$\frac{h^3 f''}{12}$	$\frac{h^5 f^{(4)}}{90}$	$\frac{3h^5 f^{(4)}}{80}$	$\frac{8h^7 f^{(6)}}{945}$	$\frac{275h^7 f^{(6)}}{12096}$	$\frac{9h^9 f^{(8)}}{1400}$	$\frac{2368h^{11} f^{(10)}}{467775}$	$\frac{1\,34635h^{13} f^{(12)}}{3269\,18592}$

一般說来,将积分区間(a, b)分为k等分,然后将各部分区間再分成n等分而使用n次的 Newton-Cotes 公式,将其結果相加作为积分的值(例如,上面的例 1 相当于$k=6$, $n=2$的情形)。这样,在計算中使用的纵坐标个数是$N=kn+1$,而点的間隔是$h=(b-a)/(kn)=(b-a)/(N-1)$。因此,例如在$n=2$的情形,将各部分区間上的誤差表示成(9.10)的形式,那么,它們的总和可以总括成

$$E=\frac{h^5}{90}\{f^{(4)}(\xi_1)+\cdots+f^{(4)}(\xi_k)\}$$

$$=\frac{h^5}{90}kf^{(4)}(\xi)=\frac{(b-a)^5}{180(N-1)^4}f^{(4)}(\xi) \qquad (9.11)$$

的形状。其中$\xi_i(i=1, \cdots, k)$是各部分区間内的某个值,而ξ是整个积分区間(a, b)内的某个值。用同样方法对于n的各种值求得的誤差如表 9.5 所示(表中略去了导数中的自变量值ξ)。

表 9.5　Newton-Cotes 公式的誤差(整个区間)

n	1	2	3	4
誤差	$\frac{(b-a)^3 f^{(2)}}{12(N-1)^2}$	$\frac{(b-a)^5 f^{(4)}}{180(N-1)^4}$	$\frac{(b-a)^5 f^{(4)}}{80(N-1)^4}$	$\frac{2(b-a)^7 f^{(6)}}{945(N-1)^6}$
n	5	6	8	10
誤差	$\frac{55(b-a)^7 f^{(6)}}{12096(N-1)^6}$	$\frac{3(b-a)^9 f^{(8)}}{2800(N-1)^8}$	$\frac{296(b-a)^{11} f^{(10)}}{467775(N-1)^{10}}$	$\frac{26927(b-a)^{13} f^{(12)}}{6538\,37184(N-1)^{12}}$

按照表 9.5,在图 9.1 中見到的傾向可以得到很好的說明。即$n=1$时(梯形法则)的誤差与$(N-1)^{-2}$成比例,$n=2$时(Simpson 法则)和$n=3$时(3/8 法则)的誤差都与$(N-1)^{-4}$成比例,但后者約为前者的$180/80=2.25$

倍等事实都相符合。

当然，上式中的 ξ 是按照不同情况而相异的。因此，上述性质并不是經常精密地成立。

还有，在表 9.3 的情形，跟着次数 n 取值 2,4,6,8,… 而逐漸增大，精确度也越来越好，这是由于 $f^{(4)}(\xi)$，$f^{(6)}(\xi)$，…不太增大的緣故。对于使这序列迅速增大的函数 $f(x)$，就可能发生提高次数反而使精确度下降的現象。設 $f(x)$ 在点 ξ 附近的 Taylor 展开式的收斂半徑为 R（在复数域内考虑，即 ξ 到 $f(x)$ 的最近奇异点的距离），当 m 很大时 $f^{(m)}(\xi)$ 約有 $m!/R^m$ 的数量級。因此，如果 $h/R=(b-a)/(N-1)R$ 不是充分小的話，即使次数增加（精确度）也不会怎样变好。（如果这一比值超过 $e\approx 3$，次数越增加反而越坏❶。）在例 1 的問題中函数 $\phi(x)$ 在有限处沒有奇异点，因而不发生这种問題。

例 2 用各种 Newton-Cotes 公式求函数 $f(x)=1/(1+x^2)$ 由 0 到 1.2 的积分而加以比較（在第 10 位作舍入的精确値是

$$\operatorname{arctg}(1.2)=0.87605\,80506).$$

表 9.6 Newton-Cotes 公式的誤差（例 2）

$N \backslash n$	1	2	3	4	5	6	Weddle	8	10	12
2	$-.030$									
3	$-.012$	$-.0^{2}59$								
4	$-.0^{2}54$	—	$-.0^{2}23$							
5	$-.0^{2}30$	$-.0^{4}49$	—	$+.0^{3}34$						
6	$-.0^{2}19$	—	—	—	$+.0^{3}19$					
7	$-.0^{2}13$	$-.0^{5}36$	$-.0^{4}19$	—	—	$-.0^{5}82$	$+.0^{5}90$			
9	$-.0^{3}76$	$-.0^{5}11$	—	$+.0^{6}21$	—	—	—	$-.0^{5}12$		
11	$-.0^{3}48$	$-.0^{6}43$	—	—	$+.0^{5}11$	—	—	—	$+.0^{6}19$	
13	$-.0^{3}34$	$-.0^{6}20$	$-.0^{5}47$	$+.0^{7}24$	—	$-.0^{6}10$	$+.0^{4}82$	—	—	$-.0^{8}86$

与例 1 中的表 9.3 和图 9.1 对应的是表 9.6 和图 9.3。

在本例中，由于 $f(x)$ 在 $x=\pm i$ 有奇异点，因此即使提高次数也不会改善精确度到象例 1 那样的程度。当使用的纵坐标的个数較少时反而会有精确度变坏的情形发生。还可以看到由表 9.5 預想的比例关系（例如 $n=1$ 时 $|E|\propto(N-1)^{-2}$）在 N 小的地方也有些遭到破坏的現象。

❶ 关于以上的討論，参看 Hildebrand [3], pp. 79～80.

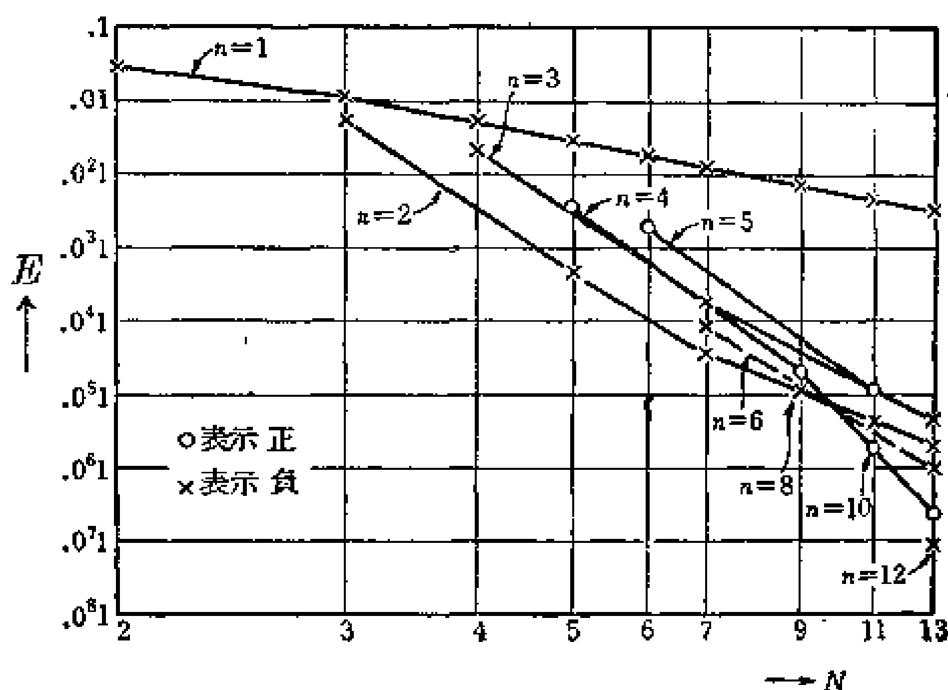


图 9.3　Newton-Cotes 公式的誤差(例 2)

§10　Чебышев 积分公式及 Gauss 积分公式

前节中討論的 Newton-Cotes 公式使用了在等分积分区間的点上所立的纵坐标，但在一般情形并不一定要这样作，因此，本节中討論使用并非等距离地，而是按别的原則适当选取纵坐标位置的积分公式。其中之一就是 Чебышев 公式。这一公式是把在积分区間内(与函数 $f(x)$ 无关地)确定的 N 个点 $x_i(i=1, \cdots, N)$ 上的纵坐标乘以相等的权 W 而后相加的結果作为积分的近似值，即

$$\int_a^b f(x)\,dx \doteq W\{f(x_1)+f(x_2)+\cdots+f(x_N)\}. \qquad (10.1)$$

特别，当 $f(x)\equiv 1$ 时，为了使(10.1)成立，必須取

$$W=(b-a)/N. \qquad (10.2)$$

自然希望能够确定 $x_1, x_2, \cdots, x_N$，使(10.1)对于任意的 N 次式成立(对于更高的次数这样的要求是不合理的)。为此，作变量替换

$$x=\frac{a+b}{2}+\frac{b-a}{2}u, \qquad (10.3)$$

选取 $u_i(i=1, \cdots, N)$ 使

$$\int_{-1}^{1} u^m du = \frac{2}{N}(u_1^m + u_2^m + \cdots + u_N^m)\ (m = 1, 2, \cdots, N) \quad (10.4)$$

成立，而后令 $x_i = \frac{a+b}{2} + u_i\frac{b-a}{2}$ 就可以了。设以这样的 $u_i(i=1, 2, \cdots, N)$ 为 N 个根的 N 次方程为

$$H_N(u) = (u-u_1)(u-u_2)\cdots(u-u_N) = 0, \quad (10.5)$$

那么

$$H_1(u) = u, \qquad H_2(u) = (3u^2-1)/3, \qquad H_3(u) = (2u^3-u)/2,$$

$$H_4(u) = (45u^4-30u^2+1)/45, \quad H_5(u) = (72u^5-60u^3+7u)/72,$$

$$H_6(u) = (105u^6-105u^4+21u^2-1)/105,$$

$$H_7(u) = (6480u^7-7560u^5+2502u^3-149u)/6480,$$

$$H_8(u) = (42525u^8-56700u^6+20790u^4-2220u^2-43)/42525,$$

$$H_9(u) = (22400u^9-33600u^7+15120u^5-2280u^3+53u)/22400,$$

$$\cdots\cdots \quad (10.6)$$

表 10.1 Чебышев 公式的横坐标

$$x_i = \frac{a+b}{2} + \frac{b-a}{2}u_i \ (i=1, \cdots, N)$$

N	u_i
2	−0.57735 02692 +0.57735 02692
3	−0.70710 67812 0 +0.70710 67812
4	−0.79465 44723 −0.18759 24741 +0.18759 24741 +0.79465 44723
5	±0.83249 74870 ±0.37454 14096 0

N	u_i
6	±0.86624 68181 ±0.42251 86538 ±0.26663 54015
7	±0.88386 17008 ±0.52965 67753 ±0.32391 18105 0
9	±0.91158 93077 ±0.60101 86554 ±0.52876 17831 ±0.16790 61842 0

它的根 $u_i(i=1, \cdots, N)$ 如表 10.1 所示[1] (在表中不包含的情形，即在 $N=8$ 和 $N\geqslant 10$ 的情形，$\Pi_N(x)$ 的根中包含着一对以上的复根，不能适合目前的要求)。又在表 10.2 中列出了誤差((10.1)的右边—左边)的表示式[2]。

表 10.2 Чебышев 公式的誤差

$$E=\frac{b-a}{N}\sum_{i=1}^{N}f(x_i)-\int_a^b f(x)\,dx$$

N	2	3	4	5
E	$-\frac{(b-a)^5 f^{(4)}}{4320}$	$-\frac{(b-a)^5 f^{(4)}}{11520}$	$-\frac{(b-a)^7 f^{(6)}}{27\,21600}$	$-\frac{13(b-a)^7 f^{(6)}}{696\,72960}$

N	6	7	9
E	$-\frac{(b-a)^9 f^{(8)}}{20321\,28000}$	$-\frac{1361(b-a)^9 f^{(8)}}{100\,32906\,24000}$	$-\frac{163(b-a)^{11} f^{(10)}}{64935\,73550\,08000}$

公式的推导 为了推导(10.6)，一般設

$$\Pi_N(u)=u^N+a_1u^{N-1}+\cdots+a_N, \tag{10.7}$$

注意到，由根与系数的关系

$$a_1=-\sum u_i,\quad a_2=\sum_{i<j}u_iu_j,\quad a_3=-\sum_{i<j<k}u_iu_ju_k,\ \cdots \tag{10.8}$$

因为它們是根的对称式，因此可以用

$$s_1=\sum u_i,\quad s_2=\sum u_i^2,\quad s_3=\sum u_i^3,\ \cdots \tag{10.9}$$

表示。但由条件(10.4)推出

$$\begin{aligned} s_m=\sum u_i^m&=0 \qquad (m=1,3,\cdots\leqslant N),\\ &=N/(m+1) \qquad (m=2,4,\cdots\leqslant N), \end{aligned} \tag{10.10}$$

这样，系数 $a_1, a_2, \cdots, a_N$ 就完全确定了。

例如，当 $N=2$ 时，$a_1=-s_1$，$a_2=(s_1^2-s_2)/2$，代以 $s_1=0$，$s_2=2/3$ 得到 $a_1=0$，$a_2=-1/3$，由此推出 $\Pi_2(u)=u^2-1/3$。(由此也推出，当 $N=2$ 时 $u_1, u_2=\pm 1/\sqrt{3}$。)

为了对于一般的 N 能够迅速地导出 $\Pi_N(u)$，下述的技巧是有用的：

---

[1] Kopal [4], p. 540; 原始文献 Salzer, Journ. Math. Phys., 26(1947), 191.

[2] Hildebrand [3], p. 350.

$$\begin{aligned}\Pi_N(u)&=\prod_{i=1}^{N}(u-u_i)=u^N\prod_{i=1}^{N}\left(1-\frac{u_i}{u}\right)=u^N\exp\left\{\log\prod_{i=1}^{N}\left(1-\frac{u_i}{u}\right)\right\}\\&=u^N\exp\left\{\sum_{i=1}^{N}\log\left(1-\frac{u_i}{u}\right)\right\}=u^N\exp\left\{\sum_{i=1}^{N}\left(-\frac{u_i}{u}-\frac{u_i^2}{2u^2}-\frac{u_i^3}{3u^3}-\cdots\right)\right\}\\&=u^N\exp\left\{-\frac{s_1}{u}-\frac{s_2}{2u^2}-\frac{s_3}{3u^3}-\cdots\right\}.\end{aligned}\tag{10.11}$$

将最后的指数函数展成 $1/u$ 的級数;用它到含 s_N 的項就可以确定 $\Pi_N(u)$，其余的項还可用于推导以 $s_1, \cdots, s_N$ 表示 $s_{N+1}, s_{N+2}, \cdots$ 的式子。

例 1 用 $N=6$ 的 Чебышев 公式求前节例 2 (55 頁) 中的积分。

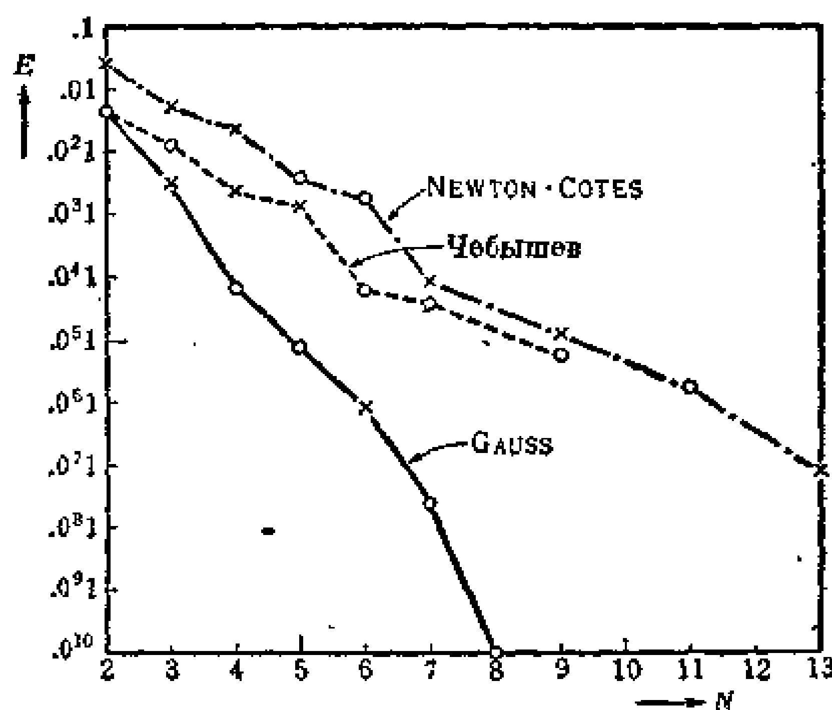


图 10.1 数值积分公式的誤差(§ 9 例 2)

表 10.3 例 1 的計算

x_i	$f(x_i)$
0.08025 19091	0.99360 08439
0.34648 88077	0.89281 37027
0.44001 87591	0.83779 00216
0.75998 12409	0.63388 56960
0.85351 11923	0.57854 25438
1.11974 80909	0.44368 80476

$0.2\sum f(x_i)=0.87606\ 41711$

由表 10.1 查出 u_i 而計算 $x_i=0.6+0.6u_i$，对于这 x_i 求 $f(x_i)$。将其总和乘以 $(b-a)/N=1.2/6=0.2$ 就是答案。記录如表 10.3，答案是 0.87606 41711。誤差是 $+0.0^5 61205$。

用同样方法以 $N=2, 3, 4, \cdots, 9$ 的 Чебышев 公式計算所得的誤差 E 如表 10.4 和图 10.1 (虚綫) 所示。

[注] 在图 10.1 中，为了便于比較，还

用鏈綫❶画出了 Newton-Cotes 公式的誤差(表 9.6 中 $N=n+1$ 时的值)。这結果很好地反映了由表 10.2 可以預想到的趋势。

表 10.4　Чебышев 公式的誤差(§ 9 例 2)

N	2	3	4	5	6	7	9
E	$0.0^{2}42$	$0.0^{2}13$	$-0.0^{3}24$	$-0.0^{3}12$	$0.0^{5}61$	$0.0^{5}39$	$0.0^{6}59$

表 10.5　Gauss 积分公式的横坐标与系数

N	u_i	A_i
2	±0.57735 02692	0.50000 00000
3	±0.77459 66692	0.27777 77778
	0	0.44444 44444
4	±0.86113 63116	0.17392 74226
	±0.33998 10436	0.32607 25774
5	±0.90617 98459	0.11846 34425
	±0.53846 93101	0.23931 43352
	0	0.28444 44444
6	±0.93246 95142	0.08566 22462
	±0.66120 93865	0.18038 07865
	±0.23861 91861	0.23395 69673
7	±0.94910 79123	0.06474 24832
	±0.74153 11856	0.13985 26957
	±0.40584 51514	0.19091 50253
	0	0.20897 95918
8	±0.96028 98565	0.05061 42681
	±0.79666 64774	0.11119 05172
	±0.52553 24099	0.15685 33229
	±0.18343 46425	0.18134 18917
9	±0.96816 02395	0.04063 71942
	±0.83603 11073	0.09032 40803
	±0.61337 14327	0.13030 53482
	±0.32425 34234	0.15617 35385
	0	0.16511 96775
10	±0.97390 65285	0.03333 56722
	±0.86506 33667	0.07472 56746
	±0.67940 95683	0.10954 31813
	±0.43339 53941	0.13463 01055
	±0.14887 43390	0.14776 21124

所謂 Gauss 积分公式，就是适当选取纵坐标的位置和权而对于一直到次数尽可能高的多項式都能精确地成立的公式。它的形状是

$$\int_a^b f(x)dx=(b-a)\{A_1f(x_1)+A_2f(x_2)+\cdots+A_Nf(x_N)\}. \tag{10.12}$$

当 $2\leqslant N\leqslant 10$ 时，可以由表 10.5 求出 u_i, A_i $(i=1, \cdots, N)$，而由

❶ 即形如 —·—·—·— 的綫。——譯者注

$$x_i=\frac{a+b}{2}+\frac{b-a}{2}u_i \quad (i=1, \cdots, N) \tag{10.13}$$

决定横坐标。

确定坐标和权的方法 表 10.5 的 u_i, A_i 是由下述的考虑决定的：作变量替换 $x=(a+b)/2+u(b-a)/2$，公式(10.12)就变成

$$\int_{-1}^{1} g(u)\,du=2\sum_{i=1}^{N} A_i g(u_i). \tag{10.14}$$

为了使这公式对于一直到次数尽可能高的多项式 $g(u)$ 精确地成立，只要能使

$$\sum_{i=1}^{N} A_i u_i^m=\frac{1}{2}\int_{-1}^{1} u^m du \tag{10.15}$$

对于 $m=0, 1, 2, \cdots$ 的一直到尽可能靠后的值成立就可以了。其中，对应于 $m=0, 1, \cdots, N-1$ 的 N 个式子，不論 $u_1, \cdots, u_N$ 取怎样的值(只要它們是相异的)，就可以适当选取系数 $A_1, \cdots, A_N$ 使它們成立。

这就相当于，令 $\Pi_N(u)=\prod_{i=1}^{N}(u-u_i)$，而决定

$$A_i=\frac{1}{2\Pi_N'(u_i)}\int_{-1}^{1}\frac{\Pi_N(u)}{u-u_i}du \quad (i=1, \cdots, N) \tag{10.16}$$

(积分 Lagrange 插补公式(8.6)就可以看到这一事实)。

这样决定系数 A_i 以后，其次考虑，如何决定坐标 u_i (或即多项式 $\Pi_N(u)$)，使(10.15)对于一直到尽可能靠后的 m 的值成立。由确定 A_i 的方法看出，对于 $m=0, 1, \cdots, N-1$, (10.15) 成立。如果假定，(10.15)对于 $m=N$ 也成立，将这些等式分别乘以 $\Pi_N(u)=u^N+a_1u^{N-1}+\cdots+a_N$ 中次数相同的項的系数而后相加就得到

$$\sum_{i=1}^{N} A_i\Pi_N(u_i)=\frac{1}{2}\int_{-1}^{1}\Pi_N(u)\,du. \tag{10.17}$$

因为每个 u_i 都是 $\Pi_N(u)$ 的根，因此左边等于 0，由此推出，右边也一定等于 0. 反之，如果右边等于 0，由(10.17)成立以及对于 $m=0, 1, \cdots, N-1$, (10.15)成立的事实推出，对于 $m=N$, (10.15)也成立。

假設 $\Pi_N(u)$ 使(10.17)的右边成为 0，因而使(10.15) 对于 $m=0, 1, \cdots$ $\cdots, N-1, N$ 成立，那么，类似地可以証明，为了使(10.15)对于 $m=N+1$ 也成立，

$$\sum_{i=1}^{N} A_i u_i\Pi_N(u_i)=\frac{1}{2}\int_{-1}^{1} u\Pi_N(u)\,du \tag{10.18}$$

的右边等于零是一个必要充分条件(不过在目前的情形要将 $m=1, 2, \cdots, N$,

$N+1$ 时的(10.15)分别乘以 a_N, a_{N-1}, …, a_1, 1 而后相加)。这样进行下去就可以看到，(10.15)对于 $m=0, 1, \cdots, N, N+1, \cdots, 2N-1$ 成立的事实与

$$\sum_{i=1}^{N} A_i u_i^m H_N(u_i) = \frac{1}{2}\int_{-1}^{1} u^m H_N(u)\,du \quad (m=0, 1, \cdots, N-1) \tag{10.19}$$

的右边都等于零的事实是等价的。换句話說，为了使(10.15)对于 $m \leqslant 2N-1$ 成立，以 $u_1, \cdots, u_N$ 为根的多項式 $H_N(u)$ 和 $N-1$ 次以下的任意多項式正交的条件是必要而充分的。但是，滿足这样条件的多項式是唯一确定的，这就是 Legendre 多項式(用 u^N 的系数除得的結果)。也就是說，取 N 次 Legendre 多項式的根为 u_i，再对于这样选取的 u_i 如上述那样决定系数 A_i 就可以了。

例 2　用 $N=4$ 时的 Gauss 公式計算 § 9 例 2 (50 頁)中的問題。

記录如表 10.6 所示，答案是 0.87607 48742，因而誤差是 $+0.0^6 68236$.

表 10.6　例 2 的計算

x_i	$f(x_i)$	A_i
0.08331 82130	0.99310 59331	0.17392 74226
0.39601 13738	0.86443 49777	0.32607 25774
0.80398 86262	0.60738 66559	0.32607 25774
1.11668 17870	0.44504 21434	0.17392 74226

$1.2\sum A_i f_i = 0.87606\ 48742$

对于 $N=2, 3, \cdots, 8$，进行类似計算的結果如表 10.7 和图 10.1 (实綫)所示。

[注]　可以看到，与使用相同个数纵坐标的其他类型的公式相比較，精确度是显著地优越的。

表 10.7　Gauss 公式的誤差(§ 9 例 2)

N	2	3	4	5	6	7	8
E	$0.0^2 42$	$-0.0^3 33$	$0.0^5 68$	$0.0^6 82$	$-0.0^7 84$	$0.0^8 28$	0.0^{10}

Gauss 积分公式的誤差一般由下式給出❶：

$$E = (b-a)\sum_{i=1}^{N} A_i f(x_i) - \int_a^b f(x)\,dx$$
$$= -\frac{(N!)^4 (b-a)^{2N+1}}{(2N+1)\{(2N)!\}^3} f^{(2N)}(\xi). \tag{10.20}$$

对于 $N=2\sim 8$，計算这一誤差的結果如表 10.8 所示。

❶　例如，參看 Kopal [4], p. 378.

表10.8 Gauss公式的誤差 E

N	2	3	4	5
E	$\frac{-(b-a)^5 f^{(4)}}{4320}$	$\frac{-(b-a)^7 f^{(6)}}{2016000}$	$\frac{-(b-a)^9 f^{(8)}}{1778112000}$	$\frac{-(b-a)^{11} f^{(10)}}{2.534876467\times 10^{12}}$

N	6	7	8
E	$\frac{-(b-a)^{13} f^{(12)}}{5.316480911\times 10^{15}}$	$\frac{-(b-a)^{15} f^{(14)}}{1.540260470\times 10^{19}}$	$\frac{-(b-a)^{17} f^{(16)}}{5.891496296\times 10^{22}}$

Чебышев积分公式和Gauss积分公式的想法和处理方法都可以毫无改变地推广到乘有变动权的积分的情形。

在Чебышев公式的情形,代替(10.1),(10.2)而令

$$\int_a^b w(x)f(x)dx = W\{f(x_1)+f(x_2)+\cdots+f(x_N)\}, \tag{10.21}$$

$$W = \int_a^b w(x)dx/N \tag{10.22}$$

就可以了。設由变换(10.3),$w(x)$变成$\bar{w}(u)$,那么,(10.4)变成

$$\frac{\int_{-1}^1 \bar{w}(u)u^m du}{\int_{-1}^1 \bar{w}(u)du} = \frac{1}{N}\sum_{i=1}^N u_i^m. \tag{10.23}$$

用(10.5)定义多項式$\Pi_N(u)$,并将它如(10.7)那样展开,那么系数(10.8)和基本对称式(10.9)之間的关系与前面完全相同。只不过代替(10.10)使用(10.23)右边的N倍就可以了。此时,(10.11)当然可以不加改变地使用。

在Gauss公式的情形,代替(10.12)令

$$\int_a^b w(x)f(x)dx = W_1 f(x_1) + W_2 f(x_2) + \cdots + W_N f(x_N), \tag{10.24}$$

設$\Pi_N(x) = (x-x_1)(x-x_2)\cdots(x-x_N)$,而由

$$W_i = \frac{1}{\Pi_N'(x_i)}\int_a^b \frac{\Pi_N(x)}{x-x_i}dx \quad (i=1, 2, \cdots, N) \tag{10.25}$$

确定 W_i, 又选取 $\Pi_N(x)$ 使它以 $w(x)$ 为权与 $x^m(m=0, 1, \cdots, N-1)$ 正交,即选取 $\Pi_N(x)$ 使

$$\int_a^b w(x)x^m\Pi_N(x)dx=0 \quad (m=0, 1, \cdots, N-1). \tag{10.26}$$

在表 10.9 中列举出几个具体的例子（表中多項式的 x^N 的系数不一定是 1, 因此, 精确地說, 应該是用 x^N 的系数除过后才是 $\Pi_N(x)$). 在最后的一例中, 系数 W_i 是常数, 因此和具有同一权 $w(x)=1/\sqrt{1-x^2}$ 的情形下的 Чебышев 型公式是一致的。

表 10.9 Gauss 型积分公式的例子

积　分	多項式 $\Pi_N(x)$	x_i, W_i 的数值表
$\int_0^\infty e^{-x}f(x)dx$	Laguérre 多項式 $L_N(x)$	Salzer-Zucker: Bull. Amer. Math. Soc., 55 (1949), 1004.
$\int_0^\infty e^{-x}x^p f(x)dx$	广义 Laguerre 多項式 $L_N^p(x)$	Burnett: Proc. Cambr. Phil. Soc., 33 (1937), 359.
$\int_0^1 f(x)\log x\, dx$	—	Mineur: Techniques de Calcul Numerique (1952), 555.
$\int_{-\infty}^\infty e^{-x^2}f(x)dx$	Hermite 多項式 $H_N(x)$	Salzer-Zucker-Capuano: N. B. S. Journ. of Res., 48 (1952), 111.
$\int_{-1}^1 \frac{f(x)}{\sqrt{1-x^2}}dx$	Чебышев 多項式 $T_N(x)$	$x_i=\cos\frac{2i-1}{2N}\pi$, $W_i=\frac{\pi}{N}$

§11 数值微分

用中心差分的記号(7.14)和求平均值的运算

$$\mu f(x)=\frac{1}{2}\left\{f\left(x+\frac{h}{2}\right)+f\left(x-\frac{h}{2}\right)\right\} \tag{11.1}$$

的記号写出 Stirling 插补公式(7.11),就得到

$$f(x)=f(x_0+hu)=f_0+u\mu\delta f_0+\frac{u^2}{2}\delta^2 f_0$$
$$+\frac{u(u^2-1)}{6}\mu\delta^3 f_0+\frac{u^2(u^2-1)}{24}\delta^4 f_0+\cdots. \tag{11.2}$$

将这表示式对 u 微分而后令 $u=0$ 就得到

$$hf'(x_0)=\mu\delta f_0-\frac{1}{6}\mu\delta^3 f_0+\frac{1}{30}\mu\delta^5 f_0-\frac{1}{140}\mu\delta^7 f_0$$

$$+\frac{1}{630}\mu\delta^9 f_0-\frac{1}{2772}\mu\delta^{11} f_0+\cdots, \tag{11.3}$$

如果对 u 微分两次而后令 $u=0$ 就得到

$$h^2 f''(x_0)=\delta^2 f_0-\frac{1}{12}\delta^4 f_0+\frac{1}{90}\delta^6 f_0-\frac{1}{560}\delta^8 f_0$$

$$+\frac{1}{3150}\delta^{10} f_0-\frac{1}{16632}\delta^{12} f_0+\cdots. \tag{11.4}$$

給定了在等距离的点上的函数值，而由此求在这些点上的导数的值时，可以利用以上的公式。

用記号推导的方法　以下用記号方法推导相当于(11.3)和(11.4)的推广的公式❶。

用

$$E^u f(x)=f(x+hu) \tag{11.5}$$

定义算子 E^u，那么(7.14)就成为

$$\delta=E^{\frac{1}{2}}-E^{-\frac{1}{2}}, \tag{11.6}$$

而(11.1)成为

$$\mu=\frac{1}{2}(E^{\frac{1}{2}}+E^{-\frac{1}{2}}). \tag{11.7}$$

另一方面，如用算子 D 表示对 x 微分的运算，那么 Taylor 展开就可以写成

$$Ef(x)=f(x+h)=f(x)+\frac{h}{1!}Df(x)+\frac{h^2}{2!}D^2 f(x)$$

$$+\frac{h^3}{3!}D^3 f(x)+\cdots=e^{hD}f(x), \tag{11.8}$$

因此

$$E=e^{hD}. \tag{11.9}$$

代入(11.6)得到

$$\delta=e^{\frac{hD}{2}}-e^{-\frac{hD}{2}}=2\,\mathrm{sh}\frac{hD}{2}, \tag{11.10}$$

由此推出

$$hD=2\,\mathrm{sh}^{-1}\left(\frac{\delta}{2}\right)=\delta-\frac{\delta^3}{24}+\frac{3\delta^5}{640}-\cdots. \tag{11.11}$$

---

❶ 参看 W. G. Bickley: J. Math. Phys., 27 (1948), 183.

这样，就用中心差分算子 δ 的級数代替了微分算子 D，但是，(11.11) 并不能直接用来求 hDf_0。这是由于，当 m 是奇数时，差分表中沒有 $\delta^m f_0$ 的值（如果是求 $hDf_{\frac{1}{2}}$，即在表中所列变量的値的中点上的导数値时是可以使用的）。但是，(11.11) 的偶次方是可以使用的。一般說来，(11.11) 的 n 次方等于

$$h^nD^n=\delta^n\Big(1-\frac{n}{24}\delta^2+\frac{5n^2+22n}{5760}\delta^4-\frac{35n^3+462n^2+1528n}{29\,03040}\delta^6+\frac{175n^4+4620n^3+4072n^2+1\,19856n}{13934\,59200}\delta^8-\frac{385n^5+16940n^4+2\,79884n^3+20\,57968n^2+56\,82048n}{36\,78732\,28800}\delta^{10}+\cdots\Big),\tag{11.12}$$

因此，当 n 是偶数时，可以用以計算 $D^nf_0=f^{(n)}(x_0)$。(11.4) 相当于在上式中令 $n=2$ 的情形。

以下想要导出类似于 (11.3) 的公式，为了这一目的，需要右边乘以 μ 的形状。但由 (11.7) 和 (11.6) 可以推出

$$\mu=\left(1+\frac{\delta^2}{4}\right)^{\frac{1}{2}},\tag{11.13}$$

因此，可以用 $\mu(1+\delta^2/4)^{-\frac{1}{2}}=1$ 乘 (11.12) 的右边而得到

$$\begin{aligned}h^nD^n&=\mu\delta^n\left(1+\frac{\delta^2}{4}\right)^{-\frac{1}{2}}\left(1-\frac{n}{24}\delta^2+\frac{5n^2+22n}{5760}\delta^4-\cdots\right)\\&=\mu\delta^n\Big(1-\frac{n+3}{24}\delta^2+\frac{5n^2+52n+135}{5760}\delta^4\\&\quad-\frac{35n^3+777n^2+5749n+14175}{29\,03040}\delta^6\\&\quad+\frac{175n^4+6720n^3+96794n^2+6\,19776n+14\,88375}{13934\,59200}\delta^8\\&\quad-\frac{385n^5+22715n^4+5\,36294n^3+63\,33250n^2+374\,08281n+884\,09475}{36\,78732\,28800}\delta^{10}\\&\quad+\cdots\Big),\end{aligned}\tag{11.14}$$

用这表示式可以計算奇数阶的导数 $D^nf_0=f^{(n)}(x_0)$。(11.3) 是在 (11.4) 中令 $n=1$ 的情形。

例 1　利用表 11.1 的差分表求 $J_0'(1.3)$。

表中在括号内記載着 $\mu\delta_0, \mu\delta_0^3, \mu\delta_0^5$。用这些値由 (11.3) 可以算得 ($h=0.1$)

$$J_0'(1.3)=\frac{10^{-5}}{0.1}\left(-5213.881-\frac{38.0445}{6}+\frac{-0.311}{30}\right)$$
$$=10^{-4}(-5220.22175-0.010367)=-0.52202\,32117.$$

表11.1 $J_0(x)$ 的差分表

x	$J_0(x)$	δ	δ^2	δ^3	δ^4	δ^5	δ^6
1.0	0.76519 769						
		−4557 567					
1.1	0.71962 202		−291 361				
		−4848 928		35 614			
1.2	0.67113 274		−255 747		1 721		
		−5104 675		37 335		−302	
1.3	0.62008 599	(−5213 881)	−218 412	($38\ 044_5$)	1 419	(−311)	−18
		−5323 087		38 754		−320	
1.4	0.56685 512		−179 658		1 099		
		−5502 745		39 853			
1.5	0.51182 767		−139 805				
		−5642 550					
1.6	0.45540 217						

[注] 在小数第8位作舍入的精确值是 $-0.52202\,325$ (依据 $J_0'(1.3)=-J_1(1.3)$ 而由 $J_1(x)$ 的表查得的)。因此，上面求得的近似值含有約为 $+0.0^74$ 的誤差。

一般說来，如果給定的函数值 f_i $(i=-3,-2,-1,0,1,2,3)$ 中含有舍入誤差 ε_i，那么，如上例那样用到 $\mu\delta_0^5$ 而求得的导数的近似值中所含的誤差为

$$\frac{1}{h}\Big[\frac{1}{2}\{(\varepsilon_1-\varepsilon_0)+(\varepsilon_0-\varepsilon_{-1})\}-\frac{1}{6}\,\frac{1}{2}\{(\varepsilon_2-3\varepsilon_1+3\varepsilon_0-\varepsilon_{-1})$$
$$+(\varepsilon_1-3\varepsilon_0+3\varepsilon_{-1}-\varepsilon_{-2})\}+\frac{1}{30}\,\frac{1}{2}\{(\varepsilon_3-5\varepsilon_2+10\varepsilon_1-10\varepsilon_0$$
$$+5\varepsilon_{-1}-\varepsilon_{-2})+(\varepsilon_2-5\varepsilon_1+10\varepsilon_0-10\varepsilon_{-1}+5\varepsilon_{-2}-\varepsilon_{-3})\}\Big]$$
$$=\frac{1}{60h}(\varepsilon_3-9\varepsilon_2+45\varepsilon_1-45\varepsilon_{-1}+9\varepsilon_{-2}-\varepsilon_{-3}). \tag{11.15}$$

如果假定，ε_i 相互独立地在区間 $(-0.5\times10^{-8},\ +0.5\times10^{-8})$ 上服从正态分布律，那么，它們的方差每个都等于 $10^{-16}/12$，因此(11.14)的方差等于

$$\frac{1}{3600\,h^2}(1+81+2025+2025+81+1)\frac{10^{-16}}{12}=\frac{2107\times10^{-16}}{21600\,h^2},$$

由此推出，(11.14)的标准离差等于

$$\sqrt{\frac{2107}{21600}}\,\frac{10^{-8}}{h}=\frac{0.31\times10^{-8}}{h}.$$

在目前的情形，由于 $h=0.1$，因此(11.15)等于 3.1×10^{-8}，可能发生它的二倍左右的誤差，也就是說，必須估計到，在小数第8位有6左右的誤差。

假定要用 Lagrange 插补公式(§8)的方法进行計算，可以依据如下的公式：

$$f'(x_0)=\frac{K'}{h}\sum A_i'f_i. \tag{11.16}$$

其中系数 A_i' 和 K' 如表11.2所示，标号 i 取表内列出的范圍內的值。

公式的推导　把(11.3)的右边在适当的地方截断，而后进行和(11.14)类似的变形就可以了。

表11.2　求导数所需的系数

i	A_i'		
	3点法	5点法	7点法
−3			−1
−2		1	9
−1	−1	−8	−45
0	0	0	0
1	1	8	45
2		−1	−9
3			1
	$K'=1/2$	$K'=1/12$	$K'=1/60$

例2　用7点法的 Lagrange 公式計算上例中的同一問題。

如表11.3所示那样进行計算，得到的答案是 −0.52202 231。

［注］　这里得到的答案，和前面在例1中得到的答案当然是相同的。

表11.3　例2的計算

x	$J_0(x)$	A'
1.0	0.76519 769	−1
1.1	0.71962 202	9
1.2	0.67113 274	−45
1.3	0.62008 599	0
1.4	0.56685 512	45
1.5	0.51182 767	−9
1.6	0.45540 217	1

$$\frac{1}{6.0}\sum A_i'f_i=-0.52202\,32317$$

表11.4　数值微分公式的誤差

$$E=(K'/h)\sum A_i'f_i-f'(x_0)$$

h \ N	3	5	7
0.3	$+0.0^{2}57$	$+0.0^{4}34$	$+0.0^{5}14$
0.2	$+0.0^{2}25$	$+0.0^{4}17$	$+0.0^{6}12$
0.1	$+0.0^{3}64$	$+0.0^{5}11$	$+0.0^{7}4$
0.05	$+0.0^{3}16$	$-0.0^{7}5$	$-0.0^{6}13$
0.04	$+0.0^{5}10$	$+0.0^{6}15$	$+0.0^{6}13$
0.03	$+0.0^{4}57$	$-0.0^{5}11$	$-0.0^{6}12$
0.02	$+0.0^{4}26$	$+0.0^{7}8$	$+0.0^{7}7$
0.01	$+0.0^{5}62$	$-0.0^{6}17$	$-0.0^{5}22$
0.005	$+0.0^{5}12$	$-0.0^{6}42$	$-0.0^{6}45$
0.002	$+0.0^{6}75$	$+0.0^{6}75$	$+0.0^{5}92$
0.001	$+0.0^{5}32$	$+0.0^{5}41$	$+0.0^{5}40$

象例1乃至例2那样，把同一个問題，对于种种不同的 h 值，乃至对于所用纵坐軸个数 N 的种种不同的值計算得的結果如表11.4和图11.1所示。

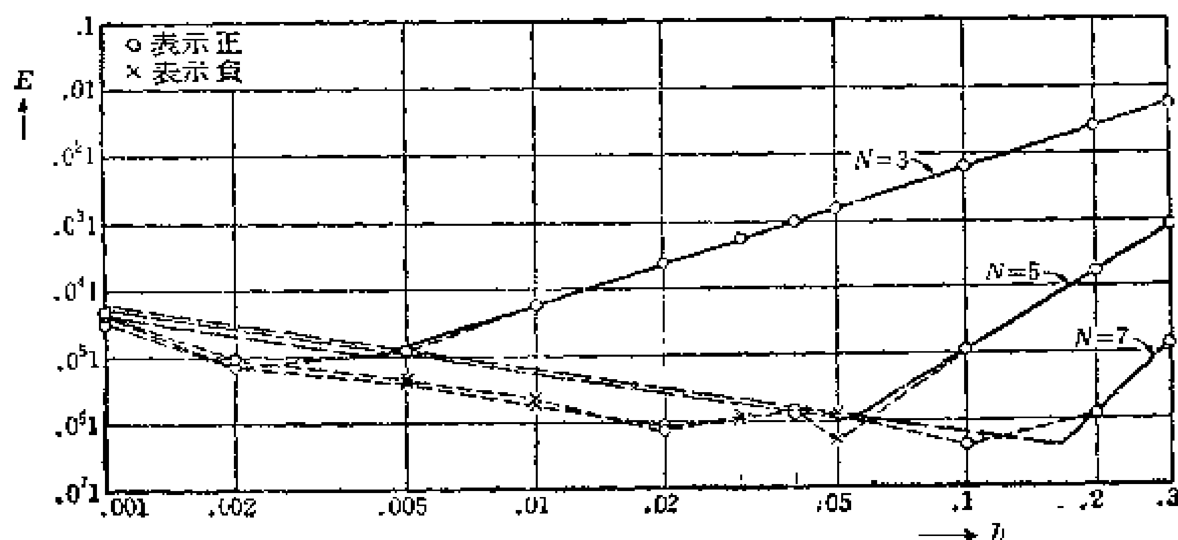


图 11.1 数値微分公式的誤差

截断誤差 E_T 的理論上的表示式如表11.5所示，因此可以說，图11.1中用实綫連接的部分的趋势，和理論是很好符合的。

表11.5 微分公式的截断誤差

N	3	5	7
E_T	$\frac{h^2}{6} f^{(3)}(\xi)$	$-\frac{h^4}{30} f^{(5)}(\xi)$	$\frac{h^6}{140} f^{(7)}(\xi)$

表11.6 微分公式的舍入誤差

N	3	5	7
σ_R	$\frac{0.20\times10^{-8}}{h}$	$\frac{0.27\times10^{-8}}{h}$	$\frac{0.31\times10^{-8}}{h}$

但是，点綫部分上的点不单和这种趋势不相符合，而且它們的排列方式也是不規則的。这当然是由于給定的函数值中所含舍入誤差的影响而产生

的。如例1的注中所述的那样計算舍入誤差的标准离差，一般地得到 $\sigma_R = K'\sqrt{\sum A_i'^2}\,10^{-8}/(\sqrt{12}\,h)$. 对于 $N=3, 5, 7$ 計算得的結果如表11.6所示。图11.1中用点綫画出它的2倍的曲綫。由此可以充分了解不規則部分的状况。

不論使用几点法，必須注意，如果把 h 分得过細，由于舍入誤差的关系，反而会使精确度下降。考虑到这种情况，在使用比之于函数值的精确度来說变数的步长相对地小的表时，有时取大于表中步长的 h 会得到更好的結果。例如，在上面的例中，对于5点法乃至7点法取 $h=0.01$ 就显得过小了，如果对于5点法取 $h=0.05$，对于7点法取0.1可能会更好一些。

第4章 函数逼近

§12 最小二乘逼近[1]

用便于計算的函数 $\varphi(x)$ 逼近某个已知函数 $f(x)$ 的要求是常常发生的。特别是，在多数情形下取 $\varphi(x)$ 为多項式。例如，取 $\varphi(x)$ 为形如 $a+bx+cx^2$ 的二次多項式，而問題在于，如何决定系数 a, b, c。 当然，决定 a, b, c 是要使 $\varphi(x)$ 与 $f(x)$ 尽可能地接近。但按照"接近"一詞的含义的不同規定，答案也是不相同的。作为一种規定方法，可以认为誤差的平方的平均值越小就越"接近"。按照这种規定，就是要决定 $\varphi(x)$ 中的系数，使

$$\int_A^B \{\varphi(x)-f(x)\}^2 dx=\min. \tag{12.1}$$

本节就处理这个問題。另一个規定方法是认为誤差絕对值的最大值越小就越"接近"，这一問題将在下一节中处理。

例1 在区間 $(0,1)$ 上用二次式 $\varphi(x)=a+bx+cx^2$ 逼近函数 $f(x)=\sqrt{1+x^2}$.

按照最小二乘逼近的原則，就是要决定 a, b, c 使

$$\int_0^1 \{(a+bx+cx^2)-\sqrt{1+x^2}\}^2 dx=\min. \tag{12.2}$$

为此，将(12.2)分别关于 a, b, c 偏微分而后令所得結果等于0：

$$a+\frac{b}{2}+\frac{c}{3}=\int_0^1\sqrt{1+x^2}dx,\quad \frac{a}{2}+\frac{b}{3}+\frac{c}{4}=\int_0^1 x\sqrt{1+x^2}dx,$$

$$\frac{a}{3}+\frac{b}{4}+\frac{c}{5}=\int_0^1 x^2\sqrt{1+x^2}dx, \tag{12.3}$$

解这方程組就可以了。右边的积分，容易由变量替换 $x=\mathrm{sh}\,u$ 求得分别为

[1] 日文"二乘"一詞的含义相当于汉文"平方"一詞的含义，因此，"最小二乘法"譯成"最小平方法"可能更为确当。但考虑到"最小二乘法"的名詞在我国沿用已久，因此在譯文中不再更改。——譯者注

$(\sqrt{2}+\mathrm{sh}^{-1}1)/2$，$(2\sqrt{2}-1)/3$，$(3\sqrt{2}-\mathrm{sh}^{-1}1)/8$。由此并使用 $\sqrt{2}=1.41421356$，$\mathrm{sh}^{-1}1=0.8813736$ 就可以定出：❶

$$a=(-33\sqrt{2}+48+3\mathrm{sh}^{-1}1)/4=0.99377,$$
$$b=(85\sqrt{2}-128+9\mathrm{sh}^{-1}1)/2=0.07026,$$
$$c=(-75\sqrt{2}+120-15\mathrm{sh}^{-1}1)/2=0.35669.$$

对种种 x 的値計算本例中得到的逼近式 $\varphi(x)=0.99377+0.07026x+0.35669x^2$，原来的函数 $f(x)=\sqrt{1+x^2}$，以及两者之差并画成图，就成为图10.1 和图 10.2 的样子。由此看来，誤差在区間(0，1)三次成为 0，而且在那里改变符号(用二次式逼近时，通常都成为这个样子)。并且，誤差的絕对値在区間的端点到达最大（在两端点都約为 0.006)，在区間内的两个地方，都取得約为 0.003 的极大値。换句話說，使用这种逼近时，在任何地方誤差也不超过 0.006，而且除了在区間端点附近很小的部分（合在一起約为全区間的 1/10 的部分)外，都在 0.003 以下。将此結果和下一节的图 13.5 比較是有意义的。

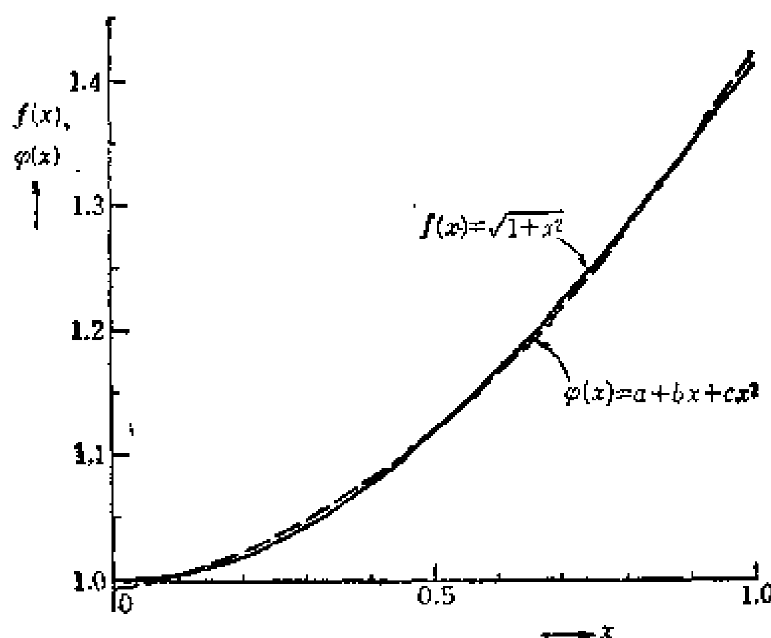


图 12.1 最小二乘逼近

例 2 在区間(0，1)上用 $\varphi(x)=cx(1-x)/2$ 逼近 $f(x)=x(1-x^2)(2-x)/24$。(参看 §7 末的注)。

---

❶ 在 Hütte 手册，24 版 (1923) 1，46 根据 Schlömilch 的結果，載有使用系数 0.9938，0.0708，0.3567 的公式。

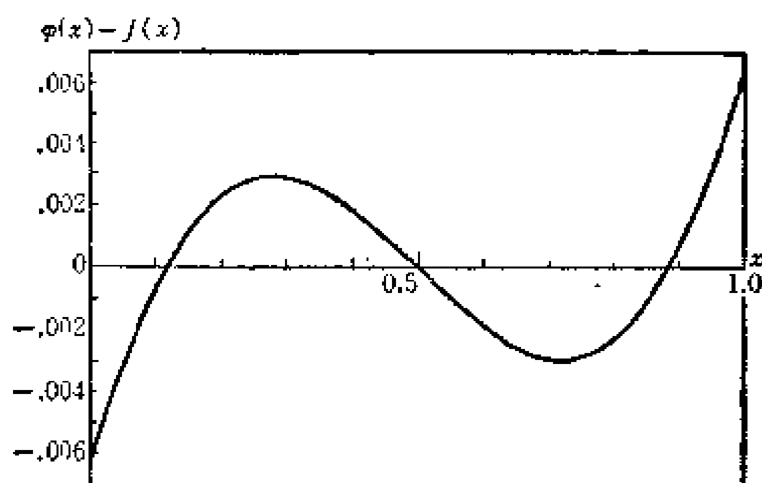


图 12.2 最小二乘逼近的誤差(例 1)

依据最小二乘法，就是要决定 c 使

$$\int_0^1 \{cx(1-x)/2 - x(1-x^2)(2-x)/24\}^2 dx = \min. \qquad (12.4)$$

为此，将上式左端对 c 微分而令所得結果等于 0，解之就得到

$$c=\int_0^1 \frac{x^2(1-x)(1-x^2)(2-x)}{48}dx \Big/ \int_0^1 \frac{x^2(1-x)^2}{4}dx = \frac{31}{168} = 0.18452.$$

本例中誤差的图如图 12.3 所示(比較后面的图 13.3)。

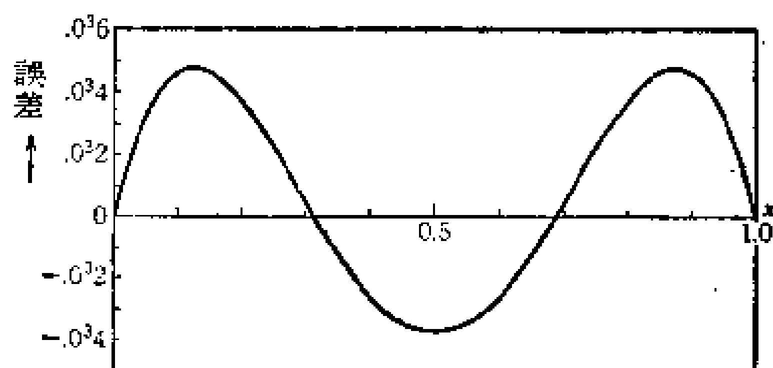


图 12.3 最小二乘逼近的誤差(例 2)

在最小二乘逼近中是使誤差平方的平均值为最小，但这"平均值"也可以是"加权平均值"。一般来說，設权为 $w(x)$，那么，就是要选取 $\varphi(x)$ 中的参数值，使

$$\int_A^B w(x)\{\varphi(x)-f(x)\}^2 dx = \min. \qquad (12.5)$$

特别是，当逼近式的形状是

$$\varphi(x)=b_0\varphi_0(x)+b_1\varphi_1(x)+\cdots+b_n\varphi_n(x), \qquad (12.6)$$

而 $\varphi_i(x)$ $(i=0, 1, \cdots)$ 在区間 (A, B) 上关于权 $w(x)$ 形成正交系，即

$$\int_A^B w(x)\varphi_i(x)\varphi_j(x)dx=0 \quad (i\neq j) \qquad (12.7)$$

时，决定系数的方法是簡单的，即

$$b_i=\int_A^B w(x)f(x)\varphi_i(x)dx\Big/\int_A^B w(x)\{\varphi_i(x)\}^2dx$$

$$(i=0, 1, 2, \cdots, n). \qquad (12.8)$$

例如，$A=-1$，$B=1$，$w(x)=1$ 时 Legendre 多項式 $P_i(x)$ $(i=0, 1, 2, \cdots)$ 形成正交系，其他的例子参看表 10.9 (64 頁)。又 $A=0$，$B=2\pi$，$w(x)=1$ 时，$\cos mx$ $(m=0, 1, 2, \cdots)$，$\sin mx$ $(m=1, 2, \cdots)$ 形成正交系。因此，容易作出形如

$$\varphi(x)=a_0+a_1\cos x+a_2\cos 2x+\cdots+a_n\cos nx$$
$$+b_1\sin x+b_2\sin 2x+\cdots+b_{n-1}\sin(n-1)x \qquad (12.9)$$

的逼近式(有限 Fourier 級数)。

此时，如同由(12.8)可以看到的那样，系数 b_i $(i=0, 1, 2, \cdots)$ 的值与 n 的值无关而只由 i 决定。即，按正交函数系的展开式，不論在那里截断，得到的都是最小二乘逼近。

到目前为止是假定函数 $f(x)$ 在区間內的所有点都已給定而进行討論的，如果 $f(x)$ 只在某些点 $x_1, x_2, \cdots, x_N$ 是給定的，那么，就应該选取 $\varphi(x)$ 中所含参变量的值，使

$$\sum_{i=1}^{N}\{\varphi(x_i)-f(x_i)\}^2=\min. \qquad (12.10)$$

此时，如果 $x_1, x_2, \cdots, x_N$ 是等距离的，而且使用关于这些点的正交多項式❶，那么，就可以得到和上述同样的結論。又当 $f(x)$ 在 $2n$ 个点 $x_k=\dfrac{k\pi}{n}$ $(k=0, 1, \cdots, 2n-1)$ 給定时，为了賦予它形如(12.9)

---

❶ 参看本丛书森口著《統計分析》§10.

的表示式，也可以使用同样的方法（为了"光滑化"，将(12.9)在中途切断而实行最小二乘逼近时系数也是相同的）。也就是，选取

$$
\left.\begin{aligned}
a_0&=\frac{1}{2n}\sum_{k=0}^{2n-1}f(x_k),\\
a_j&=\frac{1}{n}\sum_{k=0}^{2n-1}f(x_k)\cos jx_k\quad(j=1,\cdots,n-1),\\
a_n&=\frac{1}{2n}\sum_{k=0}^{2n-1}f(x_k),\\
b_j&=\frac{1}{n}\sum_{k=0}^{2n-1}f(x_k)\sin jx_k\quad(j=1,\cdots,n-1)
\end{aligned}\right\}\qquad(12.11)
$$

就可以了。这是由于，$\cos jx\ (j=0,1,\cdots,n)$，$\sin jx(j=1,\cdots,n-1)$ 关于在点組 $x_0,x_1,\cdots,x_{2n-1}$ 上求总和的运算具有正交性：

$$
\left.\begin{aligned}
&\sum_{k=0}^{2n-1}\cos jx_k\cos j'x_k=0\ (j\neq j'),\\
&\sum_{k=0}^{2n-1}\sin jx_k\sin j'x_k=0\ (j\neq j'),\\
&\sum_{k=0}^{2n-1}\cos jx_k\sin j'x_k=0\ (\text{在一切情形})。
\end{aligned}\right\}\qquad(12.12)
$$

例 3 在 12 个点 $x_k=\frac{k\pi}{6}$ $(k=0,1,\cdots,11)$ 給定 $f(x)$ 的值 $f(x_k)=f_k$ 时，賦与它以形如(12.9)的表示式（12 点法的調和分析）。

使用一般式(12.11)直接进行計算也是可以的，但如果照表 12.1 和 12.2 那样先算出数据的和与差然后进行計算可以多少节約一些时間精力。

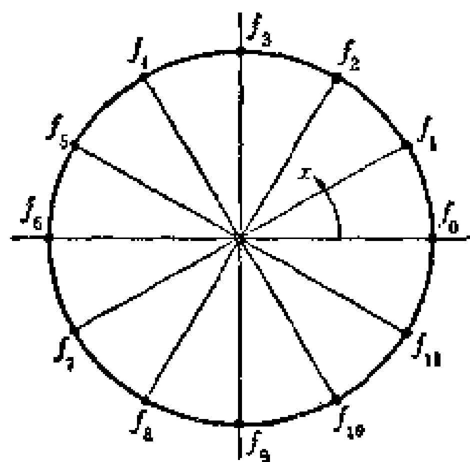


图 12.4 調和分析（12 点法）

表12.1　余弦成分

x	数据之和	$\cos x$	$\cos 2x$	$\cos 3x$	$\cos 4x$	$\cos 5x$	$\cos 6x$
0	$F_0=f_0$	1	1	1	1	1	1
$\pi/6$	$F_1=f_1+f_{11}$	$\sqrt{3}/2$	$1/2$	0	$-1/2$	$-\sqrt{3}/2$	-1
$\pi/3$	$F_2=f_2+f_{10}$	$1/2$	$-1/2$	-1	$-1/2$	$1/2$	1
$\pi/2$	$F_3=f_3+f_9$	0	-1	0	1	0	-1
$2\pi/3$	$F_4=f_4+f_8$	$-1/2$	$-1/2$	1	$-1/2$	$-1/2$	1
$5\pi/6$	$F_5=f_5+f_7$	$-\sqrt{3}/2$	$1/2$	0	$-1/2$	$\sqrt{3}/2$	-1
π	$F_6=f_6$	-1	1	-1	1	-1	1
	$12a_0$	$6a_1$	$6a_2$	$6a_3$	$6a_4$	$6a_5$	$12a_6$
	a_0	a_1	a_2	a_3	a_4	a_5	a_6

表12.2　正弦成分

x	数据之差	$\sin x$	$\sin 2x$	$\sin 3x$	$\sin 4x$	$\sin 5x$
0	$G_0=0$	0	0	0	0	0
$\pi/6$	$G_1=f_1-f_{11}$	$1/2$	$\sqrt{3}/2$	1	$\sqrt{3}/2$	$1/2$
$\pi/3$	$G_2=f_2-f_{10}$	$\sqrt{3}/2$	$\sqrt{3}/2$	0	$-\sqrt{3}/2$	$-\sqrt{3}/2$
$\pi/2$	$G_3=f_3-f_9$	1	0	-1	0	1
$2\pi/3$	$G_4=f_4-f_8$	$\sqrt{3}/2$	$-\sqrt{3}/2$	0	$\sqrt{3}/2$	$-\sqrt{3}/2$
$5\pi/6$	$G_5=f_5-f_7$	$1/2$	$-\sqrt{3}/2$	1	$-\sqrt{3}/2$	$1/2$
π	$G_6=0$	0	0	0	0	0
		$6b_1$	$6b_2$	$6b_3$	$6b_4$	$6b_5$
		b_1	b_2	b_3	b_4	b_5

§13　使最大誤差为最小的逼近

如果規定逼近式 $\varphi(x)$ 和原来的函数 $f(x)$ 之差的絕对值的上确界

$$\max_{A\leq x\leq B}|\varphi(x)-f(x)| \tag{13.1}$$

越小，$\varphi(x)$ 和 $f(x)$ 就越接近，那么，选取 $\varphi(x)$ 中所含参变量的值使(13.1)尽可能地小就可以了。这是一种极大极小化(minimax)

的方法。

例 1 用本节中的观点处理前节例 2 中的同一問題。

首先考察一下整个的趋势。对于 $f(x)=x(1-x^2)(2-x)/24$，令 $\varphi(x)=cx(1-x)/2$ 中的 c 由 0 开始逐漸增大，一开始时最大誤差（絕对值最大的誤差，下同）在区間 $(0,1)$ 的中点 $x=1/2$ 处发生，其符号是負的而大小則逐漸减小。以后（当 c 超过 $1/6=0.16667$ 后）就在左右两个点发生正的极大誤差，而其大小則逐步增大。而且，左右两处的正的极大和中間的負的极大（代数的极小）对于 c 的某个值有相等的大小。过此以后，正的极大还是逐漸增大，而負的极大則一度变成 0 后，成为正的极小而逐漸增大，最后和由左右两方向中間趋近的极大重合而成为一个极大，从此以后就一直增大下去（图 13.1）。

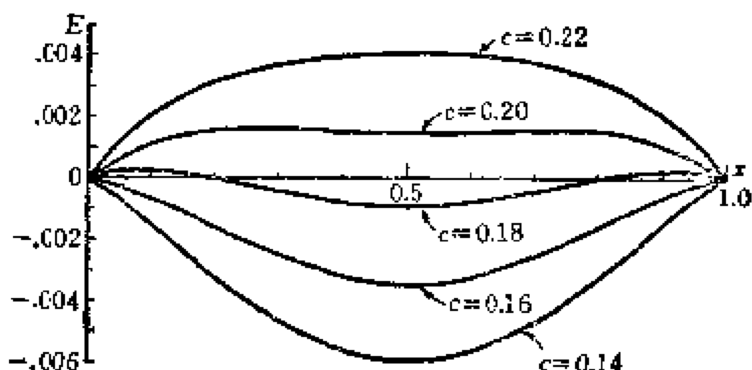


图 13.1 誤差曲綫的变化

注意到上述的情况，以下进行数式的处理，誤差是

$$E(x)=cx(1-x)/2-x(1-x^2)(2-x)/24, \tag{13.2}$$

微分之，得到

$$E'(x)=c(1-2x)/2-(1-x-3x^2+2x^3)/12$$
$$=\frac{1-2x}{2}\left(c-\frac{1+x-x^2}{6}\right). \tag{13.3}$$

当 $x=1/2$, $x=\{1\pm\sqrt{5-24c}\}/2$ 时 $E'(x)=0$，但后面的两点只在 $1/6\leqslant c\leqslant 5/24$ 时在区間 $(0,1)$ 内。在 $x=1/2$，誤差的值是

$$E(1/2)=c/8-3/128, \tag{13.4}$$

在 $x=\{1\pm\sqrt{5-24c}\}/2$，誤差的值是

$$E\left(\frac{1\pm\sqrt{5-24c}}{2}\right)=\frac{(6c-1)^2}{24}. \tag{13.5}$$

它們的图如图 13.2 所示。由此看出，$\max|E(x)|$ 当(13.4)和(13.5)大小相等而方向相反，即当

$$(6c-1)^2/24=3/128-c/8 \tag{13.6}$$

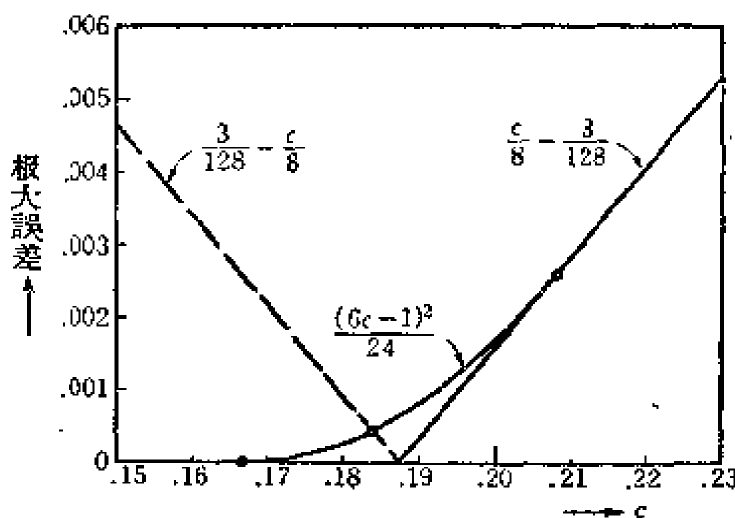


图 13.2 极大誤差的变化

时达到最小。由此解出，在 $1/6\leqslant c\leqslant 5/24$ 之間的根 $c=(3+\sqrt{2})/24\approx0.18393$。此时，誤差曲綫成为图 13.3 的形状。在 3 个地方发生最大誤差，其大小是 $(3-2\sqrt{2})/384\approx0.0004468$。

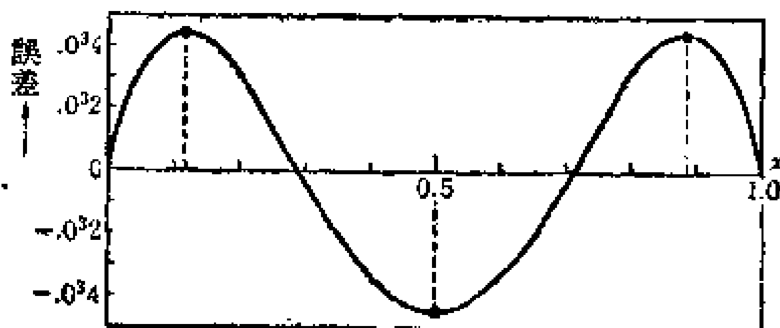


图 13.3 极大极小化逼近的誤差曲綫

这个例子稍微有些特殊。因为逼近式的形状，除了尺度因子 c 以外是确定的。如果象前节的例 1 那样，只限定了逼近式的次数，而系数则完全可以自由选取，那么，一般說来，可以作出如下的結論：

在区間 $[A,B]$ 上用 n 次式 $\varphi(x)$ 逼近給定的函数 $f(x)$ 时，如果关于某个 n 次式 $\bar{\varphi}(x)$，誤差 $\bar{E}(x)=\bar{\varphi}(x)-f(x)$ 的图形在区間 $[A,B]$ 上的 $n+2$ 个以上的点取得极值，而且这些极值的大小等于定值 M，但具有正負相間的符号，而在其余的点 $|\bar{E}(x)|<M$，那么，用任何其他 n 次式 $\varphi(x)$ 逼近 $f(x)$ 时，誤差絕对值的最大值

(13.1) 都在 M 以上。而且，为了使誤差絕对值的最大值等于 M，必須 $\varphi(x)=\bar{\varphi}(x)$。

証明 为明确起見，考慮 $n=2$ 的情形。假定对于某个二次式 $\bar{\varphi}(x)$，在区間 $[A, B]$ 上的 4 点 x_1, x_2, x_3, x_4 $(x_1<x_2<x_3<x_4)$，$\bar{E}(x)=\bar{\varphi}(x)-f(x)$ 取得极值 $-M, +M, -M, +M$（图 13.4）。于是，問題就是要証明，对于由任何其他二次式 $\varphi(x)$ 所作逼近的誤差 $E(x)=\varphi(x)-f(x)$，$\max|E(x)|>M$。

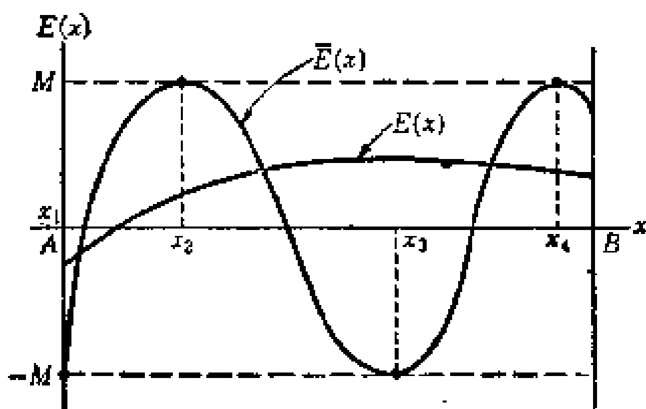


图 13.4 极大极小化逼近的特征

姑且假定，$\max|E(x)|\leqslant M$，那么，在整个区間 $[A,B]$ 上 $|E(x)|\leqslant M$，因此，当然也有 $|E(x_k)|\leqslant M$ $(k=1, 2, 3, 4)$。由此推出

$$\left.\begin{array}{ll}E(x_1)\geqslant -M=\bar{E}(x_1), & \therefore\ \varphi(x_1)-\bar{\varphi}(x_1)\geqslant 0,\\ E(x_2)\leqslant +M=\bar{E}(x_2), & \therefore\ \varphi(x_2)-\bar{\varphi}(x_2)\leqslant 0,\\ E(x_3)\geqslant -M=\bar{E}(x_3), & \therefore\ \varphi(x_3)-\bar{\varphi}(x_3)\geqslant 0,\\ E(x_4)\leqslant +M=\bar{E}(x_4), & \therefore\ \varphi(x_4)-\bar{\varphi}(x_4)\leqslant 0.\end{array}\right\} \qquad (13.7)$$

$\varphi(x)-\bar{\varphi}(x)$ 是二次式，但由 (13.7) 推出，它在区間 $[x_1, x_2]$，$[x_2, x_3]$，$[x_3, x_4]$ 上至少各有一根，总起来就至少有三个根（也可能在区間的交界上有等根，这时就把它算作是两个根）。这样的二次式只能是恒等于 0。由此可見，如果 $\max|E(x)|\leqslant M$，那就必須是 $\varphi(x)\equiv\bar{\varphi}(x)$。也就是說，只要 $\varphi(x)\neq\bar{\varphi}(x)$，就有 $\max|E(x)|>M$，只在 $\varphi(x)\equiv\bar{\varphi}(x)$ 时才有 $\max|E(x)|=M$。（証毕）。

上述性质是极大极小化逼近，即在最大誤差为最小的意义下的最佳逼近 $\varphi(x)$ 的特征性质。以下考虑寻求具有这种特征的 n 次式 $\bar{\varphi}(x)$ 的方法。

方法之一是，由任意的逼近式 $\varphi(x)=a+bx+cx^2$ 出发，計算 $E(x)=\varphi(x)-f(x)$，求出它的极值的位置 x_1, x_2, x_3, x_4 和大小 $-M_1, +M_2, -M_3, +M_4$（作为一个例子参看图 12.2）。认作最佳逼近 $\bar{\varphi}(x)=\bar{a}+\bar{b}x+\bar{c}x^2$ 的誤差 $\bar{E}(x)$ 的极值的位置和 $E(x)$ 的

极值的位置没有很大的差异，而設

$$\left.\begin{aligned}\bar{a}+\bar{b}x_1+\bar{c}x_1^2-f(x_1)&=-M,\\ \bar{a}+\bar{b}x_2+\bar{c}x_2^2-f(x_2)&=+M,\\ \bar{a}+\bar{b}x_3+\bar{c}x_3^2-f(x_3)&=-M,\\ \bar{a}+\bar{b}x_4+\bar{c}x_4^2-f(x_4)&=+M.\end{aligned}\right\}\tag{13.8}$$

将(13.8)与 $a+bx_i+cx_i^2-f(x_i)=\pm M_i$ $(i=1, 2, 3, 4)$ 边边相减，就得到关于 $\Delta a=\bar{a}-a$, $\Delta b=\bar{b}-b$, $\Delta c=\bar{c}-c$ 以及 M 的方程組

$$\left.\begin{aligned}\Delta a+x_1\Delta b+x_1^2\Delta c+M&=M_1,\\ \Delta a+x_2\Delta b+x_2^2\Delta c-M&=-M_2,\\ \Delta a+x_3\Delta b+x_3^2\Delta c+M&=M_3,\\ \Delta a+x_4\Delta b+x_4^2\Delta c-M&=-M_4.\end{aligned}\right\}\tag{13.9}$$

由此解得系数的校正量 Δa, Δb, Δc 和 $\max|\bar{\varphi}(x)-f(x)|$ 的近似值 M。当然，方程組(13.8)只不过是近似地成立，因此，目前得到的新的系数 $\bar{a}=a+\Delta a$, $\bar{b}=b+\Delta b$, $\bar{c}=c+\Delta c$ 也可能是不精确的。为此，可以把具有这些新的系数的逼近式作为新的 $\varphi(x)$ 而进行同上的处理，并反复使用这种方法直到系数的值充分收斂时为止。[1]

作为出发点的逼近二次式 $\varphi(x)$，在誤差成为0的三点，即使 $\varphi(x_i)=f(x_i)$ 的点 $x_i(i=1, 2, 3)$ 已經指定时，容易由 Lagrange 插补公式确定。(这里，x_i 的意义和前面不同，但想来不致因此引起什么混淆。) 如果使用 Чебышев 多項式 $T_3(u)\equiv 4u^3-3u=0$ 的三根 $u_1=-\sqrt{3}/2$, $u_2=0$, $u_3=\sqrt{3}/2$, 而令 $x_i=(A+B)/2+u_i(B-A)/2$，在大多数情形下可能进行得比較順利。一般說来，在 n 次式的情形，使用 $N=n+1$ 次 Чебышев 多項式 $T_N(u)$ 的 N 个根是适宜的，此时可以不必使用一般的 Lagrange 插补公式，而可以令 $u=\{x-(A+B)/2\}/\{(B-A)/2\}$，并設 $\varphi(x)$ 为

[1] Hastings: Approximation for Digital Computers (Princeton Univ. Press, 1955); A. Shenitzer, Jour. Ass. Comp. Mach., 4 (1957), 30~35.

$$\varphi(x)=a_0T_0(u)+a_1T_1(u)+\cdots+a_nT_n(u) \tag{13.10}$$

的形状，并由

$$\begin{aligned} a_0&=\frac{1}{N}\sum_{i=1}^{N}f(x_i), \\ a_k&=\frac{2}{N}\sum_{i=1}^{N}f(x_i)T_k(u_i) \quad (k=1,2,\cdots,n) \end{aligned} \tag{13.11}$$

决定系数 a_k。

公式的說明 令 $u=\cos\theta$，那么 $T_k(u)=\cos k\theta$ $(k=0,1,\cdots)$。因此，$T_N(u)=0$ 的 N 个根可以用 $\cos N\theta=0$ 的 N 个根 $\theta_i=(2i-1)\pi/2N$ $(i=1,2,\cdots,N)$ 而写成

$$u_i=\cos\theta_i=\cos\frac{(2i-1)\pi}{2N} \quad (i=1,2,\cdots,N). \tag{13.12}$$

由此也推出

$$T_k(u_i)=\cos\frac{k(2i-1)\pi}{2N}, \tag{13.13}$$

因而

$$\sum_{i=1}^{N}T_k(u_i)T_{k'}(u_i)=\begin{cases}0 & (k\neq k'), \\ N/2 & (1\leqslant k=k'\leqslant N-1), \\ N & (k=k'=0 \text{ 或 } N)\end{cases} \tag{13.14}$$

成立(参看前节的 (12.11))。

为了便于参考，在表 13.1 中列举出 $T_N(u)$ 的具体的表示式❶。

表 13.1 Чебышев 多項式

N	$T_N(u)$	N	$T_N(u)$
0	1	6	$32u^6-48u^4+18u^2-1$
1	u	7	$64u^7-112u^5+56u^3-7u$
2	$2u^2-1$	8	$128u^8-256u^6+160u^4-32u^2+1$
3	$4u^3-3u$	9	$256u^9-576u^7+432u^5-120u^3+9u$
4	$8u^4-8u^2+1$	10	$512u^{10}-1280u^8+1120u^6-400u^4+50u^2-1$
5	$16u^5-20u^3+5u$	11	$1024u^{11}-2816u^9+2816u^7-1232u^5+220u^3-11u$

❶ National Bureau of Standards, Applied Mathematics Series 9, Tables of Chebyshev Polynomials $S_n(x)$ and $C_n(x)$, 1952 中載有到 $N=12$ 的表 (对于 $T_n^*(x)=T_n(2x-1)$ 到 $n=20$)。又关于 $C_n(x)=2T_n(x/2)$ 有 $n=2(1)12$, $x=0(0.001)2$ 的 12D (到小数第 12 位)数值表。

例 2 (前节例 1). 在区間 $(0, 1)$ 上用二次式 $\varphi(x)=a+bx+cx^2$ 逼近函数 $f(x)=\sqrt{1+x^2}$.

$T_3(u)=0$ 的 3 根是 $u_1=-\sqrt{3}/2$, $u_2=0$, $u_3=\sqrt{3}/2$. 与此相对应, $x_1=1/2-(1/2)(\sqrt{3}/2)=(2-\sqrt{3})/4$, $x_2=1/2$, $x_3=1/2+(1/2)(\sqrt{3}/2)=(2+\sqrt{3})/4$. 因此

$$f(x_1)=\sqrt{1+\left(\frac{2-\sqrt{3}}{4}\right)^2}=\frac{\sqrt{23-4\sqrt{3}}}{4}=1.00224,$$

$$T_1(u_1)=-\frac{\sqrt{3}}{2},\quad T_2(u_1)=\frac{1}{2},$$

$$f(x_2)=\sqrt{1+\left(\frac{1}{2}\right)^2}=\frac{\sqrt{5}}{2}=1.11803,$$

$$T_1(u_2)=0,\quad T_2(u_2)=-1,$$

$$f(x_3)=\sqrt{1+\left(\frac{2+\sqrt{3}}{4}\right)^2}=\frac{\sqrt{23+4\sqrt{3}}}{4}=1.36767,$$

$$T_1(u_3)=\frac{\sqrt{3}}{2},\quad T_2(u_3)=\frac{1}{2},$$

而确定出系数是

$$a_0=\frac{1}{3}\{f(x_1)+f(x_2)+f(x_3)\}=1.16265,$$

$$a_1=\frac{\sqrt{3}}{3}\{-f(x_1)+f(x_3)\}=0.21098,$$

$$a_2=\frac{1}{3}\{f(x_1)-2f(x_2)+f(x_3)\}=0.04461.$$

因为 $u=(x-1/2)/(1/2)$, 因此, 这样得到的逼近式可以写成

$$\varphi(x)=1.16265+0.21098T_1(2x-1)+0.04461T_2(2x-1). \tag{13.15}$$

为了使用 81 頁脚注中的表, 写成

$$\varphi(x)=1.16265+0.10548C_1(4x-2)+0.02230C_2(4x-2) \tag{13.16}$$

的形状是便利的。如果改写成通常的形状, 那就是

$$\begin{aligned}\varphi(x)&=1.16265+0.21098(2x-1)+0.04461\{2(2x-1)^2-1\}\\&=0.99628+0.06505x+0.35691x^2.\end{aligned} \tag{13.17}$$

这样得到的 $\varphi(x)$ 的誤差的图如图 13.5 所示, 大体上滿足上述的最佳逼近条件。因此, 更进一步調整系数也沒有得到多大改进的希望。

用 n 次式逼近时, 利用 $N=n+1$ 次的 Чебышев 多項式 $T_N(u)$

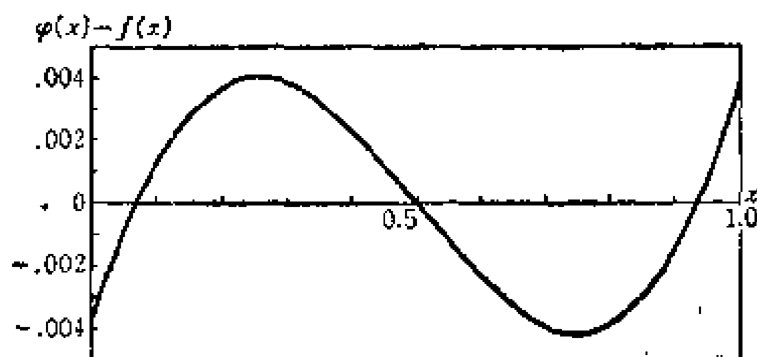


图 13.5 大体上最佳逼近的誤差（§12，例1）

的根 u_i 而作插补式(13.10)多半比較順利的原因可以說明如下：

$$f(x)=f\left(\frac{A+B}{2}+\frac{B-A}{2}u\right)=f\left(\frac{A+B}{2}+\frac{B-A}{2}\cos\theta\right) \quad (13.18)$$

是 θ 的偶函数，因此可以展开成余弦級数

$$f(x)=a_0+a_1\cos\theta+a_2\cos 2\theta+\cdots+a_n\cos n\theta+\cdots. \quad (13.19)$$

如果 $f(x)$ 是比較柔和的函数，那么，展开式的系数迅速减小，因此，将此展开式到含 $\cos n\theta$ 的項截断时的誤差大体上可以用其次的一項 $a_N\cos N\theta$ $(N=n+1)$ 来代表。但是，这一項看作 $u=\cos\theta$ 的函数不外就是 $a_NT_N(u)$，因此，在 $T_N(u)=0$ 的 N 个根 $u_i(i=1, 2, \cdots, N)$ 处誤差等于0．反之，作 n 次式 $\varphi(x)$，在 $u_i(i=1, 2, \cdots, N)$ 处和 $f(x)$ 一致，那么这 $\varphi(x)$ 应該就是(13.19)右边的 $a_0+a_1\cos\theta+\cdots+a_n\cos n\theta$ 的部分。而誤差是 $E(x)=\varphi(x)-f(x)\approx-a_N\cos N\theta=-a_NT_N(u)$，大体上满足上述的最佳逼近条件。

在表 13.2 中列举出几个为快速电子計算机使用的精密逼近式的例子。

表中在誤差的上确界一栏中記作"絕对"的表示 $\max|E(x)|$，而記作"相对"的表示 $\max|E(x)/f(x)|$．而且，不論两者中的那一个，都是使它們成为最小而选取系数的。又 No. 的一栏是表示 Hastings: Approximation for Digital Computers (Princeton Univ. Press, 1955) 中的 Sheet No.（頁数）的。

表13.2　逼近式的例子

函　数	区　間	逼　　近　　式	誤差的上确界	No.
$\sin\frac{\pi}{2}x$	$-1\leqslant x\leqslant 1$	$1.57062\,08x-0.64322\,92x^3+0.07271\,02x^5$	0.0^311 (相对)	14
$\sin\frac{\pi}{2}x$	$-1\leqslant x\leqslant 1$	$1.57079\,63184\,7x-0.64596\,37110\,6x^3$ $+0.07968\,96792\,8x^5-0.00467\,37655\,7x^7$ $+0.00015\,14841\,9x^9$	0.0^853 (相对)	16
$\sin^{-1}x$	$0\leqslant x\leqslant 1$	$\frac{\pi}{2}-\sqrt{1-x}(1.57079\,63050-0.21459\,88016x$ $+0.08897\,89874x^2-0.05017\,43046x^3$ $+0.03089\,18810x^4-0.01708\,81256x^5$ $+0.00667\,00901x^6-0.00126\,24911x^7)$	0.0^722 (絕对)	39
$\tan^{-1}x$	$-1\leqslant x\leqslant 1$	$0.99999\,93329x-0.33329\,85605x^3$ $+0.19946\,53599x^5-0.13908\,53351x^7$ $+0.09642\,00441x^9-0.05590\,98861x^{11}$ $+0.02186\,12288x^{13}-0.00405\,40580x^{15}$	0.0^737 (絕对)	13
$\log_2 x$	$\frac{1}{\sqrt{2}}\leqslant x\leqslant\sqrt{2}$	$2.88539\,00727\,38y+0.96180\,07622\,86y^3$ $+0.57658\,43420\,56y^5$ $+0.43425\,97512\,92y^7$, $y=(x-1)/(x+1)$	0.0^817 (絕对)	42
$\ln(1+x)$	$0\leqslant x\leqslant 1$	$0.99999\,64239x-0.49987\,41238x^2$ $+0.33179\,90258x^3-0.24073\,38084x^4$ $+0.16765\,40711x^5-0.09532\,93897x^6$ $+0.03608\,84937x^7-0.00645\,35442x^8$	0.0^732 (絕对)	56
10^x	$0\leqslant x\leqslant 1$	$(1+1.15129\,27760\,3x+0.66273\,08842\,9x^2$ $+0.25439\,35748\,4x^3+0.07295\,17366\,6x^4$ $+0.01742\,11198\,8x^5+0.00255\,49179\,6x^6$ $+0.00093\,26426\,7x^7)^2$	0.0^850 (相对)	20
e^{-x}	$0\leqslant x<\infty$	$1/(1+0.25001\,0936x+0.03119\,8056x^2$ $+0.00267\,3255x^3+0.00012\,7992x^4$ $+0.00001\,4876x^5)^4$	0.0^523 (絕对)	59

为了用 Taylor 展开式得到这种程度的精确度，常常需要多得多的項数。但是，Taylor 展开式的系数大多都适合簡单的規

律，因而具有可以在計算机内部算出而使用的优点。既具有这种优点，而且收斂又很快的連分数展开的应用近来也开始受到注意了❶。

［注］ Lanczos [22] pp. 346～352 提出了对于等距离的变数值給定了函数值时求 Чебышев 多項式逼近的新的方法。他的想法是，首先由被逼近函数减去适当的一次式，使所得的函数 $g(x)$ 在区間的两端点 A, B 成为 0，而后将自变量变换成 $u=\{x-(A+B)/2\}/\{(B-A)/2\}$（此时区間变成[$-1$, 1]），更进一步与 §12 例 3 类似把 $g(x)$ 表示成 $\sin k\pi u$ 和 $\cos\left(k+\frac{1}{2}\right)\pi u$ 的有限級数，而后代入

$$\sin k\pi u=2[-J_3(k\pi)T_3(u)+J_5(k\pi)T_5(u)-J_7(k\pi)T_7(u)+\cdots],$$

$$\cos\left(k+\frac{1}{2}\right)\pi u=J_0\left(\left(k+\frac{1}{2}\right)\pi\right)T_0(u)-2J_2\left(\left(k+\frac{1}{2}\right)\pi\right)T_2(u)+$$

$$+2J_4\left(\left(k+\frac{1}{2}\right)\pi\right)T_4(u)-\cdots$$

改变成 Чебышев 多項式的級数。最后的 2 段計算可以縮短成 1 段計算，准备了为此所必需的数值表（見[22]，pp. 523～528）。这一方法比之用(13.10)，(13.11)的方法，不需要求象(13.12)那样在半端点的函数值，因此是很便利的。

§14 关于数值表

作手工計算时数值表的利用价值是很高的。代替一一計算所需要的函数值，事先（把所需要的函数值）有組織地計算出来并列成表，以供查用（必要时作适宜的插补），可以說是一种常識。此时，位数和步长如何选取是一个經济性的問題，即利用价值和費用的平衡問題。1921 年在 Berlin 出版的林桂一著圓函数及双曲

❶ D. Teichroew: Math. Tables and Other Aids to Comp., 6(1952), 127～133. L. R. Shenton: Biometrika, 41(1954), 177～189; J. H. Müller: Biometrika, 22 (1930～31), 284～297; L. A. Aroian: Ann. Math. Stat., 12 (1941), 218～223; 渋谷政昭：第三回 PC セミナーテキスト，B6-3，日科技連，1957.（第三次 PC Seminar text, B6-3. 日本科技联合会，1957.）

函数的5位表共印刷了1000部，据說其中贈送給了各方面大約300部左右，而其余的經过了几年还沒有售完❶。但是后来这种数值表的需要就急驟增长，編制出来了很多龐大的函数表。虽然在使用快速电子計算机时，利用前节末所述的那样的逼近式按照需

表14.1 函数表索引(参看下页的目录)

类别	函数	文献编号
对数函数	$\log_{10}x$	3,5,2:1*,7*,8:1*,21*,26*,28:1***,28:2***.
	$\log_e x$	3,8:2,16,20,21,15*,18:2*.
	10^x	3,8:1,22:27**,28:1***.
	其他	3,21,28:1***.
指数函数	$e^{\pm x}$	6,8:2,16,20,3*,18:2*,27:1*,15**,22:14***,22:46***.
	其他	3,16,18:1,15*,18:2*,22:23***,22:41***.
三角函数	$\sin x$ 等	3,8:2,16,21,4*,10*,15*,19*,24*,18:2**,22:36**,22:40**,22:43**,22:5***.
	$\log_{10}\sin x$ 等	8:1,21,2:2*,5*,7*,26*.
	$\arcsin x$ 等	3,8:2,21,15*,18:2*,22:11**,22:26**.
	其他	3,16,20,18:2*,14:22*,15*,22:34**.
双曲函数	$\mathrm{sh}\,x$ 等	3,8:2,16,21,10*,15*,18:2*,22:36**,22:45**.
	$\log_{10}\mathrm{sh}\,x$ 等	8:1,21.
	$\mathrm{arcsh}\,x$ 等	3,8:2,21,14:20*,15*18:2*.
	其他	3,16,15*,18:2*,22:34**.
积分函数	$\mathrm{Si}(x)$ 等	3,20,14:18*,14:19*,15*,22:32*.
	$\mathrm{Ei}(x)$ 等	3,20,1*,14:21*,15*.
椭圆函数		3,15,20,18:1**,25**.
Bessel函数	$J_n(x)$	3,15,6*,17*,18:1*,11:1**,14:3***~14:14***.
	其他	3,15,6*,17*,11:1**,11:2**,18:1**,22:25**.
Gamma函数	$\Gamma(x)$, $\log_{10}\Gamma(x)$	8:1,21,3*,15*,18:2*,13:1**,22:16***.
	其他	3,23*,13:1**,13:2**.
复变函数	e^z	3.
	$\sin z$ 等	3,20,14:40* (递).
	$J_n(z),Y_n(z)$	9**,12**.
	$\Gamma(z)$	27:2*,22:34***.

編号上所附的 * 号用以指示精确度及步长的大小，沒有 * 号的4~7位，一个 * 号7~10位，** 9~14位，*** 15位以上。

❶ 林桂一：圓及双曲綫函数表、岩波，昭和16年，緒言 p. ii.

要計算到适当的程度效率更高，但用台式計算机进行計算完全成为不必要的时代看来还不会很快就来到，因此可以說，数值表的利用价值还是很大的。

以下列举手边所有的函数表中，可以认为比較重要的表的目录，并在表 14.1 中給出它的索引。

函数表目录

1 Академия Наук СССР, Таблицы Интегральной показательной Функций (1954), 300 pp.

2:1 Bauschinger und Peters, Logarithmisch-Trigonometrische Tafeln, Band 1 (1910), 360 pp.

2:2 同上, Band 2 (1910), 950 pp.

3 M. Boll, Tables Numériques Universelles (1949), 860 pp.

4 Brandenburg, Siebenstellige Trigonometrische Tafel (1931), 340 pp.

5 T. Albrecht, Bremiker's Logarithmisch-Trigonometrische Tafeln mit sechs Dezimalstellen (1924), 598 pp.

6 British Association for the Advancement of Science, Mathematical Tables, vol. 6=Bessel functions, Part I (1937), 288 pp.

7 Bruhns, A new Manual of Logarithms to Seven Places of Decimals (1934), 600 pp.

8:1 L. J. Comrie, Chamber's Six-Figure Mathematical Tables, vol. 1=Logarithmic values (1948), 580 pp.

8:2 同上, vol. 2=Natural values (1949), 570 pp.

9 Columbia University Press, Table of the Bessel Functions $J_0(z)$ and $J_1(z)$ for Complex Arguments (1947), 403 pp.

10 同上, Table of Circular and Hyperbolic Tangents and Cotangents for Radian Arguments (1947), 403 pp.

11:1 同上, Tables of Bessel Functions of Fractional Order, vol. 1 (1948), 413 pp.

11:2 同上, vol. 2 (1949), 365 pp.

12 同上, Table of the Bessel Functions $Y_0(z)$ and $Y_1(z)$ for Complex Arguments (1950), 427 pp.

13:1 H. T. Davis, Tables of the Higher Mathematical Functions, vol. 1(1933),

367 pp.

13:2 同上, vol. 2 (1935), 385 pp.

14 Harvard University, Computation Laboratory, Annals.

14:3 Tables of the Bessel Functions of the First Kind of Orders Zero and One (1947).

14:4 B. of Orders 2, 3 (1947).

14:5 B. of Orders 4, 5, 6 (1947).

14:6 B. of Orders 7, 8, 9 (1947).

14:7 B. of Orders 10, 11, 12 (1947).

14:8 B. of Orders 13, 14, 15 (1947).

14:9 B. of Orders 16～27 (1948).

14:10 B. of Orders 28～39 (1948).

14:11 B. of Orders 40～51 (1948).

14:12 B. of Orders 52～63 (1949).

14:13 B. of Orders 64～78 (1949).

14:14 B. of Orders 79～135 (1951).

14:18 Tables of Generalized Sine- and Cosine- Integral Functions, Part I (1949), 460 pp.

14:19 同上, Part II (1949), 460 pp.

14:20 Tables of Inverse Hyperbolic Functions (1949), 290 pp.

14:21 Tables of the Generalized Exponential-Integral Functions (1949), 416 pp.

14:22 Tables of the Functions $\sin\phi/\phi$ and of Its First Eleven Derivatives (1949), 241 pp.

14:40 Tables of the Function arcsinz (1956), 586 pp.

15 林桂一,高等函数表 (昭和 16 年＝1941), 222 pp.

16 林桂一,圆及双曲線函数表(昭和 16 年＝1941), 186 pp.

17 林桂一,ベッセル (Bessel) 函数表(昭和 18 年＝1943), 100 pp.

18:1 K. Hayashi(林桂一),Tafeln der Besselschen, Theta-, Kugel- und anderer Funktionen (1930),125 pp.

18:2 K. Hayashi (林桂一),Sieben und mehrstellige Tafeln der Kreis und Hyperbelfunctionen und deren Produkte sowie der Gamma Functionen (1926), 283 pp.

19 H. C. Ives, Natural Trigonometric Functions and Miscellaneous Tables (1948), 360 pp.

20 E. Jahnke und F. Emde, Funktionentafeln mit Formeln und Kurven (1933), 324 pp.

21 丸善(书店),七桁对数表(七位对数表)(昭和 28 年=1953), 504 pp.

22 National Bureau of Standards, Applied Mathematics Series (AMS).

22:5 AMS 5, Tables of Sines and Cosines to Fifteen Decimal Places at Hundredths of a Degree (1949), 95 pp.

22:11 AMS 11, Tables of Arctangents of Rational Numbers (1951), 105 pp.

22:14 AMS 14, Tables of the Exponential Functions e^x (包含 e^{-x}) (1951), 537 pp.

22:16 AMS 16, Tables of $n!$ and $\Gamma(n+1/2)$ for the First Thousand Values of n (1951), 10 pp.

22:23 AMS 23, Tables of Normal Probability Functions (1953), 344 pp.

22:25 AMS 25, Tables of the Bessel Functions $Y_0(x)$, $Y_1(x)$, $K_0(x)$, $K_1(x)$, $0 \leqslant x \leqslant 1$ (1952), 60 pp.

22:26 AMS 26, Table of Arctan x (1953), 170 pp.

22:27 AMS 27, Tables of 10^x (1953), 543 pp.

22:32 AMS 32, Table of Sine- and Cosine-Integrals for Arguments from 10 to 100 (1954), 187 pp.

22:34 AMS 34, Table of the Gamma Function for Complex Arguments (1954), 105 pp.

22:36 AMS 36, Tables of Circular and Hyperbolic Sines and Cosines for Radian Arguments (1953), 407 pp.

22:40 AMS 40, Table of Secants and Cosecants to Nine Significant Figures at Hundredths of a Degree (1954), 46 pp.

22:41 AMS 41, Tables of the Error Function and Its Derivative (1954), 302 pp.

22:43 AMS 43, Tables of Sines and Cosines for Radian Arguments (1955), 278 pp.

22:45 AMS 45, Table of Hyperbolic Sines and Cosines (1955), 81 pp.

22:46 AMS 46, Table of the Descending Exponential (1955), 76 pp.

23 K. Pearson, Tables of the Incomplet Γ-Function (1951), 164 pp.

24 J. Peters, Achtstellige Tafel der Trigonometrischen Funktionen (1939), 901 pp.

25 M. Schuler und H. Gebelein, Acht- und Neunstellige Tabellen zu den Elliptischen Funktionen, 296 pp.

26 L. Schrön, Tables des Logarithmes a Sept Décimales, 600 pp.

27:1 柴垣和三雄, 八桁函数表 (八位函数表) I, 指数函数 (昭和 24 年=1949), 104 pp.

27:2 柴垣和三雄, ガンマ函数の理論と応用 (Gamma 函数的理論及应用) (1952),

200 pp.

28:1 A. J. Thompson, Logarithmetica Britannica, vol. 1 (1952).

28:2 同上, vol. 2(1952).

第5章 常微分方程的数值解法

§15 引 論

在解决物理学上或工程中的实际問題时，如果能够把問題順利地化成数学問題，在大多数情形下就得到一个或一組微分方程。这些微分方程能用解析方法簡单地求解的情形是很少的、多半需要依靠数值方法。一提到数值解法，就有一种說不出来的、不够完整、令人厌倦的預感，因此，过去一直沒有引起数学家們多大的兴趣，大半是由不以方法本身的研究为目的的物理学家或工程师們迫于实际需要而考慮出来的。例如，所謂 Adams 方法是 Bashforth 和 Adams 为了說明毛細管現象而考慮出来的方法。但是，近来由于大型数字电子計算机的飞跃发展，上述的状况有了很大的改变，数值解法也成了专門的数学家們热烈研究的对象。不过，还沒有得到可以称为最好的結果。

一說到数值解法，可能就会产生一种近似方法的感覺，但是，通常在把問題数学化时（即把实际問題化成数学問題时），首先就加上了很多假設条件，而且还在作了很大的近似代換以后才用解析的方法求解的，这与尽可能地减少近似表示而直接用数值方法求解，究竟那一方而更接近于实际，实在是值得考慮的。数值解法是更为直接的、最适用于实际問題的、而且只要在函数的連續性或者滿足 Lipschitz 条件等比較寬广的条件下就可以使用的强有力的方法。

在开始叙述解法之前，这里先提出几点注意。首先，在施行数值解法之前，还是要在不改变問題实质的程度以内尽可能地予以簡单化、无因次化以后，能用解析的方法求解的就用解析的方法求

解，而且，尽可能地确定出可以确定的奇异点等的位置。为此，或者画出等傾斜綫，或者画出所求函数值和它的一阶导数的值所构成的拓扑曲面而考察大概的状况。然后，用变量替换等方法除去可以除去的奇异点。其次，在数值解法中，微分，积分等极限运算通常要用所謂差分的有限运算来代替，因此，必須根据要求的精确度适当选定步长的大小。最后，还要根据所用計算机的性能选用适当的解法。

§16 問題的类型与解法的类型

解常微分方程的問題大体上可以分成两种类型。一类是初值問題，在这类問題中，进行数值解法所必要的在最初几个点的初始条件都是已經給定的；另一类是边值問題，在这类問題中，在积分区間两端点上已經給定了边界条件。在前一类問題中，立即就可以进行积分，但在后一类問題中，在最初几个点的已知条件是不够充分的，因此，如不附加必要的假設条件，就不能够立即进行积分。关于这一类問題将在 §24 中作簡单的介紹，这里主要叙述初值問題。而且，以后主要将就如下的一阶微分方程进行討論：

$$y'=f(x, y) \quad a\leqslant x\leqslant b, \tag{16.1}$$

其中假定，給定了 y 在 $x=a\equiv x_0$ 的值 y_0。

［注］ 以下叙述的形如(16.1)的微分方程的解法，对于一阶微分方程組几乎可以毫无变更地适用，而高阶微分方程經过适当的变形可以化成一阶微分方程組。

用数值方法解 (16.1) 主要使用差分法，也就是把給定的积分区間 $a\leqslant x\leqslant b$ 分割为长度等于 $h=(b-a)/N$ 的子区間（N 是子区間的个数），而后如上面所說的那样，用差分的有限运算代替微分的极限运算。此外，有时还用解析的方法如 Picard 方法，Taylor 展开，解析开拓等作为輔助工具。这些方法用以求在初始点 x_0 附

近的值是很便利的。

如果把依賴于差分法的解法加以分类，大体上有如下的几种类型：

1. 前进型　卽一直到 $x=nh=x_n$ 时的 y 的值 y_n 为已知，而求下一个 y 的值 y_{n+1} 时，不需要 y'_{n+1} 的类型❶

2. 迭代型　为了求 y_{n+1}，首先用某种方法預定出 y_{n+1} 或 y'_{n+1} 的值，而后迭次代入公式加以改善，直到相邻的两个近似值相一致为止。

3. 反饋型　前两种类型的結合，由前进型預定 y_{n+1}，而由迭代型予以校正，卽具有反饋作用 (feed back) 的类型。

这些方法的优缺点将在以后說明，以下先按照求数值解时必要的順序說明这些方法中的某些方法。

§ 17 最初几个值的求法

在以差分法为基础的公式中，为了进行积分，除了在 x_0 的值以外，通常还需要知道在 x_0 附近若干个点上积分 y 的值。本节叙述积分法中便于求这些値的方法。

1) Taylor 展开法　在 $f(x, y)$ 具有比較簡单的形状时用这种方法是比較順利的（如果 $f(x, y)$ 中含有平方根，或者具有其他較为复杂的形状时，用这种方法就不很便利）。

为簡单計，設 $x_0=0$，在它的附近将 y 展开就得到

$$y=y_0+\frac{x}{1!}y'_0+\frac{x^2}{2!}y''_0+\frac{x^3}{3!}y'''_0+\cdots. \tag{17.1}$$

这一展开式只要利用 y 在 $x=0$ 的值 y_0 和原来的微分方程(16.1)就可以求得。

［注］ (17.1) 中的各級导数的值 y'_0，y''_0 等可以如下求得：

---

❶ 記号 y'_n 表示 $(y')_{x=x_n}=f(x_n, y_n)$. 有时也写作 f_n.

$$y'=f, \quad y''=f_x+ff_y, \quad y'''=f_{xx}+2ff_{xy}+f^2f_{yy}+y''f_y, \cdots$$

例 1 $y'=x-y^2$, $y(0)=0$。

由此算得

$$y''=1-2yy', \quad y'''=-2(y')^2-2yy'', \quad y^{\text{IV}}=-6y'y''-2yy''',$$
$$y^{\text{V}}=-6(y'')^2-8y'y'''-2yy^{\text{IV}}, \cdots.$$

代入 $x=0$ 时 $y=0$ 就得到如下的展开式:

$$y=\frac{1}{2!}x^2-\frac{6}{5!}x^5+\frac{252}{8!}x^8-\cdots.$$

設步长为 $h=0.1$, 就得到在开首 3 点的 y 的值:

$$y_1=y(0.1)=0.0049995,$$
$$y_2=y(0.2)=0.0199840,$$
$$y_3=y(0.3)=0.0448789.$$

y_3 的最后一位有等于 1 的舍入誤差。

在这方法中,可用以級数收斂范圍內的点为中心的 Taylor 展开式进行解析开拓而逐步得到积分的値,但在大多数情形下計算是很麻煩的。在类似于本例的情形,即 $f(x, y)$ 較为簡单而且收斂性很好,常常用这方法来求开始的几个値。在上例中,如果令 $y(0)=1$, 就已經不簡单了。

2) **Picard 方法**(迭代法) 在 Picard 方法中,将(16.1)改写成

$$y=y_0+\int_{x_0}^{x} f(x, y)\,dx \tag{17.2}$$

的形状,将右边的 y 代以第 1 个推測値 $y=y^0(x)$, 其中 $y^0(x)$ 是 x 的某个适当的已知函数,将右边积分,設这样得到的値为 $y^{(1)}$:

$$y^{(1)}=y_0+\int_{x_0}^{x} f(x, y^0(x))\,dx. \tag{17.3}$$

再将这个 $y^{(1)}(x)$ 代入 (17.2) 的右边,而令积分的結果为 $y^{(2)}(x)$。反复使用这一方法就得到

$$y^{(k+1)}=y_0+\int_{x_0}^{x} f(x, y^{(k)}(x))\,dx. \tag{17.4}$$

这样逐步得到的 $y^{(k)}(x)$ $(k=0, 1, 2, \cdots)$ 如果收敛的話，它的极限就給出精确的解。这一方法就是当 $f(x, y)$ 在积分范圍內連續、有界，而且满足 Lipschitz 条件时解的 Cauchy 存在定理証明方法的直接应用(参看本丛书《常微分方程》, p. 33)。

例 2 $y'=1-y$, $y(0)=0$.

在本例中，公式(17.4)的形状是

$$y^{(k+1)}=\int_0^x (1-y^{(k)})dx.$$

由 $y^{(0)}=x$ 开始进行計算。

$$y^{(1)}=\int_0^x (1-x)\,dx=x-\frac{x^2}{2},$$

$$y^{(2)}=\int_0^x\left(1-x+\frac{x^2}{2}\right)dx=x-\frac{x^2}{2}+\frac{x^3}{3\cdot 2},$$

$$\cdots\cdots,$$

$$y=x-\frac{x^2}{2}+\frac{x^3}{3\cdot 2}-\frac{x^4}{4\cdot 3\cdot 2}+\cdots=1-e^{-x}.$$

[注] 在本例中，求得了解析形状的解。不过，(17.4) 的右边能象本例中这样积分出来的情形是很稀少的，多数情形在 $k=2$ 左右积分已經不是很容易的了。例如，

$$y'=y^2, \quad y(0)=1,$$

令 $y^{(0)}=1$，那么 $y^{(1)}=1+x$，其次，

$$y^{(2)}=\frac{2}{3}+\frac{(1+x)^3}{3}=1+x+x^2+\frac{x^3}{3},$$

$$y^{(3)}=\frac{55}{63}+\frac{4}{9}x+\frac{1}{9}(1+x)^4+\frac{1}{63}(1+x)^7$$

$$=1+x+x^2+x^3+\frac{2}{3}x^4+\frac{1}{3}x^5+\frac{x^6}{9}+\frac{x^7}{63}.$$

所求的解实际上是 $(1-x)^{-1}=1+x+x^2+\cdots$, $|x|<1$，但如果象这样計算下去，看来实在是很不容易的。

又如，对于 $y'=\sin x+\cos y$ 进行同样的計算，立刻就会遇到困难。

但是，如果把这积分用前面讲过的数值积分进行計算，就可以避免上述的困难，而成为非常有效的方法。也就是說，把(17.4)的

被积函数 $f(x, y^{(k)})$ 一般説来用 n 次多项式近似表示，将此多项式代入(17.4)而后依次由 x_0 到 $x_1, x_2, \cdots$ 积分就得到所求的解。例如，在 x 变化时 $f(x, y^{(k)})$ 所画的曲綫上取3点(或5点)，用通过这些点的2次(或4次) Lagrange 插补公式近似表示 $f(x, y^{(k)})$ 而后积分，就可以分别得到如下的公式❶：

3点近似公式：

$$\left.\begin{aligned} y_1-y_0&=\frac{h}{12}(5f_0+8f_1-f_2)+\frac{h^4}{24}y^{\mathrm{IV}}(s),\\ y_2-y_0&=\frac{h}{3}(f_0+4f_1+f_2)-\frac{h^5}{90}y^{\mathrm{V}}(s),\ x_0<s<x_2. \end{aligned}\right\}\quad(17.5)$$

5点近似公式：

$$\left.\begin{aligned} y_1-y_0&=\frac{h}{720}(251f_0+646f_1-264f_2\\ &\quad+106f_3-19f_4)+\frac{27h^6y^{\mathrm{VI}}(s)}{1440},\\ y_2-y_0&=\frac{h}{90}(29f_0+124f_1+24f_2\\ &\quad+4f_3-f_4)+\frac{16h^6y^{\mathrm{VI}}(s)}{1440},\\ y_3-y_0&=\frac{h}{80}(27f_0+102f_1+72f_2\\ &\quad+42f_3-3f_4)+\frac{27h^6y^{\mathrm{VI}}(s)}{1440},\\ y_4-y_0&=\frac{4h}{90}(7f_0+32f_1+12f_2\\ &\quad+32f_3+7f_4)-\frac{8h^7y^{\mathrm{VII}}(s)}{945}. \end{aligned}\right\}\quad(17.6)$$

---

❶ 导出这些公式的方法可以参看柴垣[17]pp.49～65，Milne[14]pp.42～43。这些公式也可以把 Newton 插补公式

$$f(x)=f_0+u\Delta f_0+\frac{u(u-1)}{2!}\Delta^2f_0+\frac{u(u-1)(u-2)}{3!}\Delta^3f_0+\cdots$$

由0到1，2，… 积分而得到(参看 §9，p.49)。(17.5)的第二式相当于 Simpson 法則(9.1)。也有4点近似和6点近似公式，可以参看 Milne (同上) p.48，49。

各式的最后一项，就是所謂截断誤差。

应用这些公式求最初几个值的方法可以依据如下的程序：

程序 1 以后将要叙述的前进型中比較簡单的类型，例如，由

$$y_1=y_0+hf_0,\quad y_i=y_{i-2}+2hf_{i-1}\quad (i=2,3,4)\tag{17.7}$$

計算第 1 个推測值 $y_i^{(0)}$ $(i=1,2,3,4)$ 就得到 $f_i^{(0)}$.

程序 2 把上面得到的 $f_1, f_2, \cdots$ 和 f_0 代入 (17.5) 或 (17.6) 就得到 y 的第 1 近似值。在多数情形下它和第 0 近似值是不相同的。将此值代入微分方程的右边而求 $f_1, f_2, \cdots$.

程序 3 把这些 $f_1, f_2, \cdots$ 和 f_0 再代入 (17.5) 或 (17.6) 而計算 y. 反复使用程序 2, 3 一直到求得所需要的位数为止。

这种操作方法如用图解表示，就成为图 17.1 的样子。这种方法适用于机械計算，例如可以利用簡单的120位 UNIVAC 来进行。

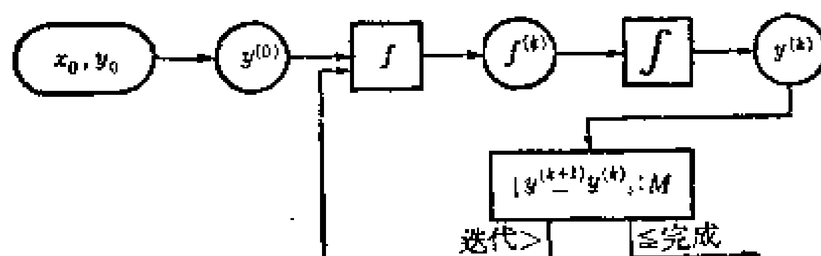


图 17.1 Picard 方法的流动图

f: 微分方程的右边， $\int$: (17.5) 或 (17.6)， $y^{(0)}$: 第 0 近似值，
M: 为判定程序 3 所需的小的正数。

例 3 $y'=(y-x)/(y+x),\quad y(0)=1.$

按 (17.7) 算出第 0 近似值以后，用 (17.6) 算得的結果如表17.1(a)所示。(第 0 近似值不一定要計算到象表 17.1 所示的那样詳細。)

[注] 作为程序 1，还有种种别的方法。例如，由于 $y'(0)=1$，取步长 $h=0.05$，就得到第 0 近似值 $y_1=1.05$, $y_2=1.10$, $y_3=1.15$, $y_4=1.20$. 由这些值开始进行計算也可以，或者令 $y'=(1-x)/(1+x)$，就得 $y=\log|1+x|-x=1+x-x^2+\frac{2x^3}{3}$. 由此得到， $y_1=1.048$, $y_2=1.091$, $y_3=1.130$, $y_4=1.165$. 由这些值开始也可以。

表17.1　$y'=(y-x)/(y+x)$，$y(0)=1$ 的最初几个值

(a) $h=0.05$

x	$y^{(0)}$	$f^{(0)}$	$y^{(1)}$	$f^{(1)}$	$y^{(2)}$	$f^{(2)}$	$y^{(3)}$
0	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
0.05	1.050000	0.909091	1.047762	0.908906	1.047654	0.908897	1.047653
0.10	1.090909	0.832061	1.091182	0.832100	1.091128	0.832092	1.091127
0.15	1.133206	0.766178	1.131195	0.765844	1.131039	0.765815	1.131037
0.20	1.167527	0.634377	1.166742	0.707333	1.167841	0.707568	1.167842

(b) $h=0.1$

x	$y^{(0)}$		$y^{(3)}$	$f^{(3)}$	$y^{(4)}$
0	1.000		1.000000	1.000000	1.000000
0.1	1.100		1.091146	0.832094	1.091146
0.2	1.167	……	1.167855	0.707571	1.167855
0.3	1.241		1.233508	0.608740	1.233507
0.4	1.289		1.290153	0.526670	1.290152

用 (17.6) 中的截断誤差来估計,这一結果的誤差就得在小数第5位的2单位以下(誤差)。(高阶导数可以写成 $(y+x)y'=y-x$ 的形状后逐次微分而求得。) 如果最初几个值的精确度还不够充分,可以把 h 取得更小一些。

即使取較大的 h, 例如取步长为 0.1, 第4近似值在小数第6位也已經收斂了。(在表17.1(b)中記录了第3近似值和第4近似值。) 但其截断誤差約为小数第4位的1单位左右。

在开始計算之前,首先根据在积分范圍內所需要的精确度,对最初几个值要求的精确度也就确定了。在子区間的数目較多因而积分的次数很大时，最初几个值的精确度应比所需要的精确度多取1~2位(当然也要考虑到函数的性质)。在积分次数較少的情形,可以多取一位或者取相同的位数就可以了。最初几个值的誤差在以后的計算过程中也有逐漸扩大的时候，也有并不这样的时候,但一般說来,应該作为逐漸扩大的情况而加以注意。在图17.2

中画出了誤差迅速增长的情形(图 a) 和增长較慢的情形(图 b) 的略图(参看第 113 頁)。

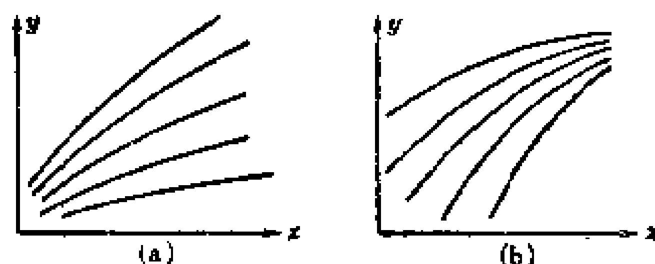


图 17.2

§18 前进型解法

属于这一类型的首先有，由 (x, y) 平面上已知点，例如点 (x_n, y_n) 处的斜率 (Euler 方法) 或在 x_n, x_{n+1} 之間的点处的斜率 (Runge 乃至 Runge-Kutta 方法等)求新的点 (x_{n+1}, y_{n+1}) 等类型。这些方法除了給定的初始值以外,不需要另外求最初的值(或者需要的話，也只要求一个左右)，就可以逐次进行积分。在步长一定的情况下，Runge 方法的計算比較簡单,但精确度却不够好。Runge-Kutta 方法的精确度是相当好的，但計算却比較繁杂。而且,不論那个方法,都无法判断計算过程中所得到的值究竟精确到什么程度,因而不得不另外檢查誤差,而且也不易发現計算过程中的錯誤。

此外,在編制完差分表的同时就得到所求解的 Adams 方法也属于这种类型,但这方法需要先求出最初的几个值。

1) **Euler 方法**　这一方法最原始的形式是将 y_{n+1} 在点 x_n 展开至含 y' 的項:

$$y_{n+1}=y_n+hy'_n+\frac{h^2}{2}y''(s),\quad x_n<s<x_{n+1}. \tag{18.1}$$

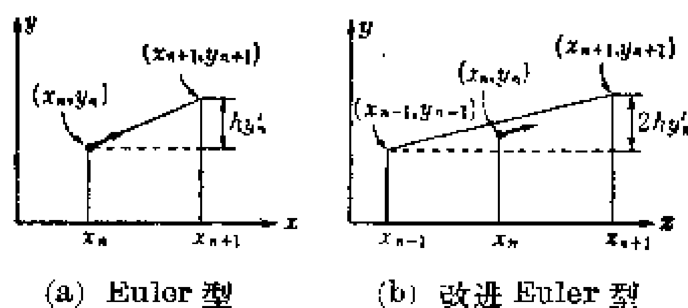
右边的第 3 項是截断誤差。这一方法可以看作是以折綫的极限証明微分方程解的存在性的 Cauchy-Lipschitz 方法为基础。(18.1)

虽然简单，但截断誤差很大，因此，代替(18.1)常使用以下的形式:

$$y_{n+1}=y_{n-1}+2hy'_n+\frac{h^3}{3}y'''(s),\ x_{n-1}<s<x_{n+1}. \qquad (18.2)$$ [1]

右边的第3項是截断誤差。这公式的精确度比原始形式的 Euler 公式要好。把 y_{n+1}, y_{n-1} 在点 x_n 展开而后加以組合就可以得到这一公式。虽然除初始值外还要計算一个最初的值，但只要作和(18.1)类似的計算就可以得到，因此常用于求后面将要叙述的反饋法的預告值。只用这种方法計算下去是很不可靠的，关于这一問題，将在§22中結合数值的例子加以說明。

以上两方法的几何意义如图18.1所示。



(a) Euler 型　　(b) 改进 Euler 型

图 18.1

2) **Runge-Kutta 方法**　这方法是历来常用的一个方法，它是把在点 x_n, $x_{n+1}=x_n+h$, 以及中間点 $x_{n+1/2}=x_n+\frac{h}{2}$ 等处的几个斜率算出，而后予以加权平均的方法。其中，最常用的是下列的公式:

$$\left.\begin{aligned}&y_{n+1}=y_n+(k_1+2k_2+2k_3+k_4)/6,\\&k_1=hf(x_n,\ y_n),\\&k_2=hf(x_{n+1/2},\ y_n+k_1/2),\\&k_3=hf(x_{n+1/2},\ y_n+k_2/2),\\&k_4=hf(x_{n+1},\ y_n+k_3).\end{aligned}\right\} \qquad (18.3)$$

[1] 在本书中把它姑且叫作改进 Euler 方法。

k_i ($i=1, 2, 3, 4$) 的意义，可以由图 18.2 明显地看出。这公式既不需要预先計算最初的几个值，精确度也比較好(截断誤差为 $O(h^5)$)，而且也容易变更步长，但是，它的缺点是，計算稍現复杂，而且必須另外估計誤差。

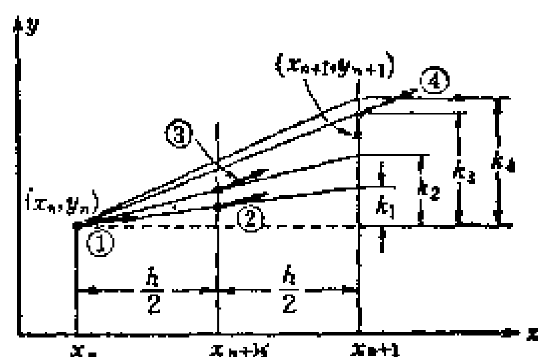


图 18.2

例 1 $y'=3y(1+x)^{-1}$, $y(0)=1$. 設 $h=0.1$, 求 y 至 $x=1.0$.

用 Runge-Kutta 方法求得的結果如表 18.1 所示。至 $x=0.2$ 記录了詳細的計算，其余只記录了計算的結果。

表 18.1 $y'=3y(1+x)^{-1}$; $y(0)=1$, $h=0.1$, $x=0\sim1.0$

(Runge-Kutta 方法)

x	y	y'	$k_i=hf$	$k=(k_1+2k_2+2k_3+k_4)/6$	E	$(1+x)^3$
0	1.00000	3.00000	0.300000	—	0	1.00000
0.05	1.15000	3.28571	0.328571	0.330982	—	—
0.05	1.16429	3.32654	0.332654	—	—	—
0.10	1.33265	3.63450	0.363450	—	—	—
0.10	1.33098	3.62995	0.362995	—	2	1.33100
0.15	1.51248	3.94560	0.394560	0.396983	—	—
0.15	1.52826	3.98677	0.398677	—	—	—
0.20	1.72966	4.32415	0.432415	—	—	—
0.20	1.72796	—	—	—	4	1.72800
0.30	2.19694	—	—	—	6	2.19700
0.40	2.74391	—	—	—	9	2.74400
0.50	3.37488	—	—	—	12	3.37500
0.60	4.09583	—	—	—	17	4.09600
0.70	4.91279	—	—	—	21	4.91300
0.80	5.83174	—	—	—	26	5.83200
0.90	6.85869	—	—	—	31	6.85900
1.00	7.99964	—	—	—	36	8.00000

［注］ 誤差（E 栏）逐漸增大。缺点在于沒有校正这种誤差的方法。事实上，作者在实际計算本例时，中途曾发生了計算上的錯誤，而且好不容易才得发現。計算一个阶段(step)所需要的时間約10分钟，熟练了的話可以稍快一些。

Euler 方法乃至 Runge-Kutta 方法多半可以用 602A 穿孔計算机或 UNIVAC-120 进行計算（参看森口 [21] p. 57）。

此外，这些方法也可以用来計算最初的几个值。

§19 积分的进行方法

虽然只用上一节所述的方法也可以进行积分，但以下还介紹两种反饋型的方法。这些方法虽兼用前进型和迭代型两类方法，但其用法在基本想法上与单独使用这些方法是不相同的。它是用前进型的公式（以下用 P 表示[1]）預告其次的一个值（上一节的方法就是只由这种类型組成的），然后把这个值用迭代型的公式（以下用 C 表示[2]）予以校正的方法。一般的程序如下（参看图 19.1）：

程序1　算出必要的最初几个值（用 §17 的方法）。

程序2　用預告算子 P 算出 y_{n+1} 的預告值，用 $y_{n+1}^{(0)}$ 表示此值。

程序3　代入給定微分方程的右边（在图 19.1 中用 f 表示）而計算 $y_{n+1}^{(0)\prime}(=f(x_{n+1}, y_{n+1}^{(0)}))$.

程序4　利用 $y_{n+1}^{(0)\prime}$ 由校正算子 C 算出 $y_{n+1}^{(0)}$ 的校正值，用 $y_{n+1}^{(1)}$ 表示这一值。

程序5　如果 $y_{n+1}^{(0)}$ 和 $y_{n+1}^{(1)}$ 接近到所希望的程度，也就是說，如果

$$|C_0|=|y_{n+1}^{(1)}-y_{n+1}^{(0)}|\leqslant M,\quad M \text{ 为常数}, \tag{19.1}$$

那么，就可以认为这个阶段的計算已經完成，于是进入下一个阶段

[1] P=predictor（預告算子）。

[2] C=corrector（校正算子）。

的計算。

程序 6 如果 $y_{n+1}^{(0)}$ 和 $y_{n+1}^{(1)}$ 沒有接近到所希望的程度（$|C_0|>M$），那么就用 $y_{n+1}^{(1)}$ 迭次使用程序 3 至 5 一直到相邻的两个校正值一致到所希望的程度。

这些方法对于机械計算也是很适宜的。以下叙述其中精确度不太高但計算比較簡单的一种方法和精确度稍高但計算稍为复杂的一种方法。

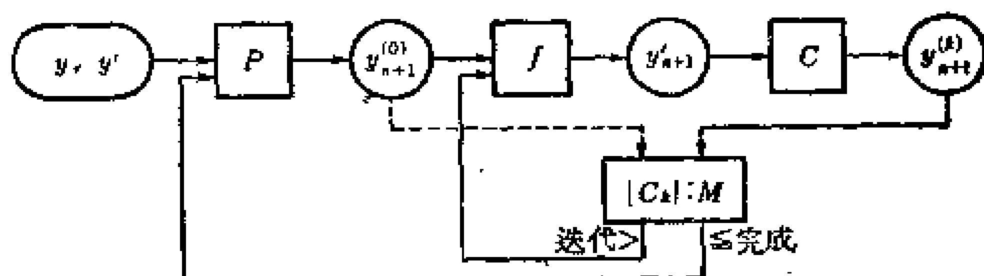


图 19.1 反饋型的流动图

y, y'：預告算子中所需要的到 x_n 为止的已知值。 P：預告算子。
C：校正算子。 f：微分方程的右边。

在$\left\{\begin{matrix}第1次循环中\\第k+1次循环中\end{matrix}\right\}$收斂性由$\left\{\begin{matrix}C_0=y_{n+1}^{(1)}-y_{n+1}^{(0)}\\C_k=y_{n+1}^{(k+1)}-y_{n+1}^{(k)}\end{matrix}\right\}$判定。

1) **梯形法則** 用于精确度的要求不太高的情形；但是，如果 h 取得很小，也可以提高精确度到某种程度。除初始值外，需要再算出一个最初的值。

$$\left.\begin{aligned}&\text{預告算子：}\ y_{n+1}=y_{n-1}+2hy'_n+h^3y'''(s_1)/3,\\&\text{校正算子：}\ y_{n+1}=y_n+(h/2)(y'_n+y'_{n+1})\\&\qquad\qquad\qquad -h^3y'''(s_2)/12,\end{aligned}\right\}\quad(19.2)$$ ❶

其中 $x_{n-1}<s_1<x_{n+1}$，$x_n<s_2<x_{n+1}$。各式右边的第 3 项都是代表截断誤差的，以下分别用 T_P 和 T_C 表示。如果 y''' 是連續的，那么

$$C_0=T_C-T_P=-\frac{5}{12}h^3y'''(s),\quad x_{n-1}<s<x_{n+1}.$$

---

❶ Heun 就把 Euler 公式(18.1)用作 P，改进形式可能是 Milne 首先采用的。这一方法所以称为梯形法則，是由于它的校正算子的形式。

如果 y^{IV} 是連續而且有界的（$|y^{\mathrm{IV}}|<L$），那么

$$|T_C-C_0/5|=|h^3(y'''(s_2)-y'''(s))/12|$$
$$=|h^3y^{\mathrm{IV}}(s')(s_2-s)/12|<h^3L|s_2-s|/12<h^4L/6.$$

其中 s' 位于 s_2 与 s 之間。由此推出，如果和 h^4 属于同等数量級的項可以略去不計的話，下列的关系式成立：

$$T_C=C_0/5. \tag{19.3}$$

由此看出，如果不考虑舍入誤差，那么可以由 C_0 估計截断誤差。此外，如果在計算过程中，把 C_0 的值記录下来，那么，如果它的值突然变得很大，就可以看作是計算发生錯誤的信号，此时，需要对計算进行檢查。由以上两点可以看出 C_0 的重要性。

例1 求 $y'=(y-x)/(y+x)$，$y(0)=1$ 在 $x=0\sim1$ 之間的解，精确到小数第3位。因为 $y''(0)=8$，因此，如果取 $h=0.1$，大体上就可以了。用 Taylor 級数求得 $y_1=1.0912$，由此就可以进行計算，結果如表 19.1 所示。表中只記录了最后結果和 C_0 的值，而且都是一次計算完了的。如果 C_0 过小的話，可以把 h 增大，反之，如果 C（校正）需要計算两次以上的話，可以把 h 縮小。

表 19.1 $y'=(y-x)/(y+x)$，$y(0)=1$，$x=0\sim1.0$，$h=0.1$ 梯形法則

x	y	y'	C_0
0	1.000	1.0000	0
0.1	1.0912	0.8321	0
0.2	1.168	0.7076	2
0.3	1.234	0.6086	1
0.4	1.291	0.5269	1
0.5	1.340	0.4565	1
0.6	1.383	0.3949	1
0.7	1.420	0.3396	1
0.8	1.451	0.2892	0
0.9	1.478	0.2431	0
1.0	1.500	0.2000	0

对于象这种类型的(比較簡单的)$f(x, y)$用 UNIVAC-120 就很可以了。即使用手工計算,也不会化費很多的时間。

以下考虑在这一类迭代法中解收斂于一定值的条件。假設由于 y_{n+1} 有等于 C_0 的差异而引起的 y'_{n+1} 的校正为 C'_0, 那么

$$C'_0=f_y C_0, \quad f_y \equiv \partial f/\partial y.$$

因此,在下一步計算中施加于 y_{n+1} 的校正是

$$C_1=hC'_0/2=(hf_y/2)C_0.$$

迭次进行計算得到

$$C_k=(hf_y/2)^k C_0 \quad (k=1, 2, \cdots),$$

因此,这种迭次計算的結果收斂的一个必要条件是

$$|hf_y/2|<1. \tag{19.4}$$

显然,当 f_y 的值很大时必須取較小的 h 的值。实际上,最好取 h 的值充分小使 C_1 到达可以忽略不計的程度。

2) **Milne 方法** 用于 h 的值不太小而希望得到較高精确度的情形,初始值之外还需要 3 个最初的值。

預告算子:

$$y_{n+1}=y_{n-3}+\frac{4h}{3}(2y'_{n-2}-y'_{n-1}+2y'_n)+\frac{28}{90}h^5y^{\mathrm{V}},$$

校正算子:

$$y_{n+1}=y_{n-1}+\frac{h}{3}(y'_{n-1}+4y'_n+y'_{n+1})-\frac{1}{90}h^5y^{\mathrm{V}}. \tag{19.5}$$

如果使用便于迭代的中心差分記号,还可以写成

預告算子:

$$y_{n+1}=y_{n-3}+4hy'_{n-1}+\frac{8h}{3}\delta^2y'_{n-1}+\frac{28}{90}h^5y^{\mathrm{V}},$$

校正算子:

$$y_{n+1}=y_{n-1}+2hy'_n+\frac{h}{3}\delta^2y'_n-\frac{1}{90}h^5y^{\mathrm{V}}. \tag{19.5'}$$

其中的校正算子也就是(17.5)的第 2 式。預告算子可以将 y_{n+1} 和

y_{n-3} 在点 x_{n-1} 展开而求得。

例 2 求 $y'=x-y^2$, $y(0)=1$ 在 $x=0\sim1.0$ 之間的解，精确到小数第5位。依据 Milne 方法在点 $x=0$ 求 y^{V} 就得到 $y^{V}(0)=-186$. 設 $h=0.1$, 那么，$T=10^{-5}(186)/90\approx2\cdot10^{-5}$. 精确度稍显不足。求最初几个値精确到小数第6位是比較适宜的。利用 $y^{VI}(0)=1192$ 而用公式 (17.6) 求这些值，这里，$T\approx22.45h^6$, 因此，需要取 $h=0.05$. 最初的几个値計算到第6近似值而在求到小数第7位后作了舍入。(与第5近似值至多在小数第7位差3. 用台式計算机計算約需1.5小时。) 最后結果如表 19.2(a) 所示。由于 C_0 逐漸减小，对于 $x=1.0$, 另外設 $h=0.1$ 而計算出的結果，如表 19.2(b) 所示，与 $h=0.05$ 的情形完全一致，而且 $|C_0|=8$. 計算每一阶段用台式計算机約需 6～8 分钟时間。此外，在記录过程中如用鉛笔記录預告値，而用鋼笔記录校正值，那么中途的笔誤可以减少，計算完后只要把鉛笔記录擦掉就可以了。

截断誤差的估計和收斂性的确定与梯形法則的情形相同，因此只举示結果如下：收斂条件是

$$|hf_y/3|<1. \tag{19.6}$$

由校正算子和預告算子的差 C_0 得到，截断誤差 T_C 的估計为

$$T_C=C_0/29. \tag{19.7}$$

此外，关于累积誤差 (p. 113) 的估計等，可以参看 Milne [14] p. 66～69。

这个方法虽然被称为 Milne 方法，但和柴垣方法几乎是一样的，只不过預告算子稍有不同而已。Milne 方法适宜于机械計算 (但使用 UNIVAC-120 类型的机器可能稍难一些)。与向来使用的 Runge-Kutta 方法相比較有很多优异之点。前面用 Runge-Kutta 方法計算过的 §18 例1 如用 Milne 方法計算，就成为表 19.3 的样子。如用 (17.6) 計算最初的几个值，到第6近似值就成为精确的值。用 Milne 方法时，由預告算子已經得到精确的值(由解析解的形状可以看出，这是很自然的)。但是，用 Runge-Kutta

表 19.2 (a) $y'=x-y^2$, $y(0)=1$, $h=0.05$

x	y	y'	$\frac{\delta^2 y'}{3}$	C_0
0	1.000000	−1.000000	—	—
0.05	0.953592	−0.859337	−0.005449	—
0.10	0.913794	−0.735020	−0.004487	—
—	—	—	(−0.003744)	—
0.15	0.879866	−0.624164	−0.003740	—
(0.20)	(0.851200)	(−0.524541)	(−0.003158)	—
0.20	0.851191	−0.524527	−0.003154	−9
(0.25)	(0.827264)	(−0.434347)	(−0.002696)	—
0.25	0.827255	−0.434351	−0.002699	−9
(0.30)	(0.807628)	(−0.352263)	(−0.002319)	—
0.30	0.807621	−0.352252	−0.002318	−7
0.35	0.791914	−0.277128	−0.002032	−2
0.40	0.779806	−0.208097	−0.001799	−7
0.45	0.771014	−0.144463	−0.001608	−8
0.50	0.765279	−0.085652	−0.001459	−2
0.55	0.762376	−0.031217	−0.001333	−2
0.60	0.762090	−0.019219	−0.001237	−2
0.65	0.764236	+0.065943	−0.001151	−2
0.70	0.768626	+0.109214	−0.001089	−1
0.75	0.775102	+0.149217	−0.001029	−2
0.80	0.783497	+0.186132	−0.000984	−1
0.85	0.793666	+0.220094	−0.000940	−2
0.90	0.805459	+0.251236	−0.000907	−1
0.95	0.818745	+0.279657	−0.000868	+1
1.00	0.833382	+0.305474	—	—

(b) $y=x-y^2$, $y(0)=1$, $x=1.0$, $h=0.1$

x	y	y'	$\frac{\delta^2 y'}{3}$	C_0
0.60	0.762090	+0.019219	—	—
0.70	0.768626	+0.109214	−0.004359	—
0.80	0.783497	+0.186132	−0.003938	—
0.90	0.805459	+0.251236	(−0.003626)	—
(1.00)	(0.833390)	(+0.305461)	—	—
1.00	0.833382	—	—	−8

方法时，如表18.1所示的那样，誤差逐漸增大，而且沒有校正的方法。

此外，还有同种类的精确度更好的公式，但計算也随之更为复杂。不过在某些問題中，下节叙述的方法1有时会比較好一些。

表19.3　$y'=\frac{3y}{1+x}$, $y(0)=1$, $h=0.1$

x	y_i	y'_i	$\frac{\delta^2 y'}{3}$	C_0
0	1.00000	3.00000	—	⋮
0.1	1.33100	3.63000	0.02000	⋮
0.2	1.72800	4.32000	0.02000	⋮
0.3	2.19700	5.07000	0.02000	⋮
(0.4)	(2.74400)	(5.88000)	(0.02000)	⋮
0.4	2.74400	5.88000	0.02000	0
0.5	3.37500	6.75000	—	—

［注］　在 Milne 方法中，不用(17.6)求最初的三个值，而把开始几点处的步长减小，用梯形法則也可以。

在积分区間不太大，而子区間的个数較少的情形用下述的方法也可以(Kunz [8] p. 200)。

程序1　用公式

$$y'_1=y'_0+hy''_0,\quad y'_{-1}=y'_0-hy''_0$$

預告y'_1和y'_{-1}。

程序2　利用上面得到的值由公式

$$\left.\begin{aligned}y_1&=y_0+(h/24)(7y'_0+16y'_1+y'_{-1})+h^2y''_0/4-h^5y^{\mathrm{V}}(s_1)/180,\\ y_{-1}&=y_0-(h/24)(y'_1+16y'_0+7y'_{-1})+h^2y''_0/4+h^5y^{\mathrm{V}}(s_2)/180\end{aligned}\right\}\qquad(19.8)$$

施行逐代法直至y_1, y_{-1}都各自一致时为止。

程序3　得到y_1, y_{-1}后由公式

$$y_2=y_0+(2h/3)(5y'_1-y'_0-y'_{-1})-2h^2y''_0+7h^5y^{\mathrm{V}}(s_3)/45\qquad(19.9)$$

預告y_2。

程序4　用 Milne 方法(19.5)或(19.5')的校正算子求y_2的校正值。

这样就得到了y_{-1}, y_0, y_1, y_2, 以下就可以轉向 Milne 的方法。

§20 其 他 解 法

在本节中，叙述与 Milne 方法有同等程度的精确度，或虽具有更高的精确度、但容易用于存储量比较少的計算机的几个方法。

1) **Obrechkoff 公式** (Milne [14] p. 76)

$$
\left.\begin{aligned}
&\text{預告算子:}\ y_{n+1}=y_{n-2}+3(y_n-y_{n-1})+h^2(y''_n-y''_{n-1})\\
&\qquad +60h^5y^{\mathrm{V}}(s_1)/720,\\
&\text{校正算子:}\ y_{n+1}=y_n+(h/2)(y'_{n+1}+y'_n)\\
&\qquad -(h^2/12)(y''_{n+1}-y''_n)+h^5y^{\mathrm{V}}(s_2)/720,\\
&\quad x_{n-2}<s_1<x_{n+1},\quad x_n<s_2<x_{n+1}.
\end{aligned}\right\}\quad(20.1)
$$

在开始的两个阶段中，把

$$y_1=y_0+hy'_0+h^2y''/2 \qquad (20.2)$$

用作預告算子并使用(20.1)中的校正算子就可以了。象 Milne 方法那样，求最初的几个值不是很困难的，但需要求 y 的二阶导数。其他的步驟与 Milne 方法相同。

2) **跳越法** (1) (Clippinger-Dimsdale 方法。Kunz [8] p. 206) 这方法与 Milne 方法不同，預告算子的精确度不好，而以迭代为主，但一次就可以同时求得两个值。首先，由 y_0 求出 y'_0，然后由

$$\text{預告算子:}\qquad y_2=y_0+2hy'_0 \qquad (20.3)$$

預算出 y_2，由此并求得 y'_2. 其次，由

第一校正算子:

$$
\left.\begin{aligned}
&y_1=\frac{1}{2}(y_0+y_2)+(h/4)(y'_0-y'_2)-h^4y^{\mathrm{IV}}(s_1)/24,\\
&\qquad x_0<s_1<x_2
\end{aligned}\right\}\quad(20.4)
$$

算出 y_1. 由此求出 y'_1，而用与 Milne 方法(19.5)相同的校正算子作为

第二校正算子：

$$\left.\begin{aligned} &y_2=y_0+(h/3)(y_0'+4y_1'+y_2')-h^5y^{\mathrm{V}}(s_2)/90, \\ &\qquad x_0<s_2<x_2, \end{aligned}\right\} \quad (20.5)$$

由此可以校正 y_2 的值。以此为基础校正 y_1 的值，而后进行迭代直到一致到所希望的位数为止。

因为預告算子的精确度不好，因此迭代的次数相应地增多（是这方法的缺点）；但只要知道了 y_n 就可以进行計算，而且一次就可以进行两个阶段，精确度也不太坏（是它的优点）。

3) **跳越法** (II) 这方法是把方法 (I) 的精确度提高了的方法，原理方法与 (I) 完全相同。

$$\left.\begin{aligned} &\text{預告算子：} y_2=y_0+2hy_0'+2h^2y_0'', \\ &\text{第一校正算子：} \\ &\quad y_1=\frac{1}{2}(y_0+y_2)+\frac{5h}{16}(y_0'-y_2')+\frac{h^2}{16}(y_0''+y_2'') \\ &\qquad -h^6y^{\mathrm{VI}}(s_1)/720, \\ &\text{第二校正算子：} \\ &\quad y_2=y_0+\frac{16h}{15}y_1'+\frac{7h}{15}(y_0'+y_2')+\frac{h^2}{15}(y_0''-y_2'') \\ &\qquad +h^7y^{\mathrm{VII}}(s_2)/4725. \end{aligned}\right\} \quad (20.6)$$

与 (I) 类似，迭代的次数也比較多。

以上三个方法，都适宜于記忆装置少的計算机，例如，如用 UNIVAC-120，上面的計算在 1) 和 2) 的情形約有 15 个，在 3) 的情形約有 20 个程序阶段 (program-step) 就可以了，其余的都可以用于計算 y' 和 y''。

§21 誤差与最适步长[1]

关于这个問題作一般的討論是比較困难的，因此，以下以梯形

[1] 关于本节第一部分所討論的誤差产生問題請参看柴垣著作[文献17]。

法則为一个简单的典型加以考虑。

設 $y'=f(x, y)$ 在点 x_n 的精确解是 y_n，那么

$$y_n=y_{n-1}+\int_{x_{n-1}}^{x_n} f(x, y(x))\,dx. \tag{21.1}$$

在实际計算时，右边被积函数中的 $y(x)$ 应当用逐次近似值代入。現在假定用充分收斂的值 Y_n（所謂充分收斂是指第 k_n 次和第 k_n+1 次代入所得的值包含舍入誤差在內达到一致）作为数値解，那么

$$e_n=Y_n-y_n \tag{21.2}$$

是数値解在点 x_n 的誤差。

假定用梯形法則进行計算，那么 (21.1) 成为

$$\left.\begin{aligned} &y_n=y_{n-1}+\frac{h}{2}(f_{n-1}+f_n)-T_n,\\ &T_n=(h^3/12)f''(s), \quad x_{n-1}\leqslant s\leqslant x_n. \end{aligned}\right\} \tag{21.3}$$

其中 $f_n=f(x_n, y_n)$.

依据 (21.3)，在第 k_n+1 次代入中得到的值 $Y_n^{k_n+1}$ 是

$$Y_n^{k_n+1}=Y_{n-1}+\frac{h}{2}(F_{n-1}+F_n). \tag{21.4}$$

其中 $F_{n-1}\equiv f(x_{n-1}, Y_{n-1})$，$F_n\equiv f(x_n, Y_n)$. 这一 $Y_n^{k_n+1}$ 是舍入以前的值，实际上在作右边的計算时需要四舍五入，因而产生舍入誤差。舍入后所得的結果就是 Y_n（Y_n 就是这样定义的）。因此，舍入誤差 $R_n^{k_n}$ 等于

$$R_n^{k_n}=Y_n-Y_n^{k_n+1}. \tag{21.5}$$

$R_n^{k_n}$ 的值依不同的計算方法而异，但是，可以考虑例如下述的两种情形：

(i) 在求 hF_n 时舍去小数第 $r+1$ 位，而在求 $\frac{1}{2}(hF_{n-1}+hF_n)$ 时至小数第 r 位四舍五入，此时

$$Y_n^{k_n+1}+R_n^{k_n}=Y_{n-1}+\frac{1}{2}[(hF_{n-1}+\varepsilon_{n-1})+(hF_n+\varepsilon_n)]+\varepsilon_n'.$$
(21.6)

与(21.4)比較得到

$$R_n^{k_n}=\frac{1}{2}(\varepsilon_{n-1}+\varepsilon_n)+\varepsilon_n'. \quad (21.7)$$

如設 $|\varepsilon_n|\leqslant 10^{-(r+1)}$, $|\varepsilon_n'|<0.5\times 10^{-r}$, 那么 $|R_n^{k_n}|\leqslant 0.6\times 10^{-r}$.

(ii) 計算 hF_n 时取足够的位数(至少取至 $r+2$ 位),用 2 除时四舍五入,此时

$$R_n^{k_n}=\varepsilon_n', \quad |\varepsilon_n'|\leqslant 0.5\times 10^{-r}.$$

为了使討論簡化,以下考慮(ii)的情形。

这样,舍入誤差的产生就弄清楚了。

其次,将(21.2)改写成

$$e_n=(Y_n-Y_n^{k_n+1})+(Y_n^{k_n+1}-y_n) \quad (21.8)$$

而加以考慮,第1項就是上述的舍入誤差 $R_n^{k_n}$,而第2項是舍入以前的第 k_n+1 个近似値对精确値 y_n 而言的誤差。由(21.3),(21.4)得到

$$Y_n^{k_n+1}-y_n=(Y_{n-1}-y_{n-1})+\frac{1}{2}h[(F_{n-1}-f_{n-1})+(F_n-f_n)]+T_n. \quad (21.9)$$

应用中値定理将上式右边[ ]內的( )改写而仍考慮(21.8),就得到

$$e_n=e_{n-1}+\frac{1}{2}h(g_{n-1}e_{n-1}+g_ne_n)+T_n+R_n.$$

其中 $g_n=(\partial f/\partial y)_n$ 是 $\partial f/\partial y$ 在 (x_n, η_n) 的値而 η_n 是 y_n 和 Y_n 之間的某个値。又 $R_n^{k_n}=R_n$. 将上式改写成

$$e_n=(1+hg_{n-1}/2)(1-hg_n/2)^{-1}e_{n-1}+T_n(1-hg_n/2)^{-1}+R_n(1-hg_n/2)^{-1}, \quad (21.10)$$

它給出第 $n-1$ 阶段和第 n 阶段的誤差的关系。由(21.10)可以看

到，各阶段的誤差有3种。右边的第1項是前面各阶段中已經存在的誤差的影响，即在前面各阶段产生的誤差的累积，因此叫做累积誤差（也叫做遺傳誤差 inherited error）。第2項是截断誤差（truncation error），即由于用有限的差商代换連續的微商而产生的（数值积分公式所具有的）誤差❶。第3項是把得到的值在所要的位数进行舍入而产生的舍入誤差（round-off error）。如果 e_{n-1} 的系数的絕对值比1小，那么，它的影响逐渐减弱，如果比1大，那么就会逐渐扩大。在一般情形，若就 $g_n=(\partial f/\partial y)_n$ 的符号和值沒有急驟变化的情形加以考虑，則当 $(\partial f/\partial y)<0$ 时誤差逐渐縮小，而在 $(\partial f/\partial y)>0$ 时誤差逐渐扩大。上面的討論是图17.2的稍較詳細的定性解釋。

誤差 e_n 满足的差分方程(21.10)是綫性的，因此 e_n 可以表示成只有截断誤差 T_n 时(21.10)的解 t_n 和只有舍入誤差 R_n 时的解 r_n 的和。因此，以下首先在 $\partial f/\partial y=g$（定数）和 $T_n=T$（定数）的假定条件下試求只有截断誤差时(21.10)的解 t_n。（当 g_n 和 T_n 的变化不太激烈时，假定它們为定数而加以处理可以了解大概的傾向。）

t_n 满足的微分方程是

$$t_n=(1+hg/2)(1-hg/2)^{-1}t_{n-1}+T(1-hg/2)^{-1}. \quad (21.11)$$

設 $t_0=0$ 而求它的解就得到

$$t_n=(T/hg)(q^n-1),\ q=(1+hg/2)(1-hg/2)^{-1}. \quad (21.12)❷$$

其次，对于只有舍入誤差 R_n 时的解 r_n，設 R_n 所能取的最大值为 R，那么和上面类似地可以求得 r_n 的最大值为

$$\max(r_n)=(R/hg)(q^n-1). \quad (21.13)$$

---

❶ 由(21.3)所确定的 T_n 表示由数值积分公式的解减去精确解后所得的值，但有时也把截断誤差定义作符号与此相反的值。

❷ 由(19.4)推出，q 不能是負的。

这一估計对于象制作数值表时那样要求較高精确度的場合是适宜的，但对于通常的計算也許就太精密了。在通常的情形，把 R_n 看作是以相同的概率取 $-R$ 到 $+R$ 之間的一切值的随机变量可能更实际一些。当計算的阶段数很大时，r_n 可能取的值的分布可以认为是接近于以0为中心而具有下列标准高差的正态分布❶：

$$D(r_n)=R(q^{2n}-1)^{\frac{1}{2}}(6hg)^{-\frac{1}{2}}. \qquad (21.14)$$

由此推出，$|r_n|$ 小于 (21.14) 的 $D(r_n)$ 的概率为 68.3%，而小于 $2D(r_n)$ 的概率为 95.4%.

以下具体考察一下，将积分区間 L 分成 N 等分（即令 $h=L/N$），而由 $x=0$ 开始积分到 $x=L$ 时在数值解中可能产生的誤差。

例　考虑如下的例子：

$$y'=1-y, \quad y(0)=0, \quad x=0\sim1.0. \qquad (21.15)$$

因为 $\partial f/\partial y=g=-1$，因此属于 $g<0$ 的情形。一般說来，当 $g<0$ 时，設 $G=-g$ 就可以将 q^N 改写成

$$q^N=\{(1-hG/2)/(1+hG/2)\}^N\approx e^{-LG}(1+O(h^2)),$$

在本例中，$q^N\approx e^{-1}=0.368$. 由此推出，

$$\left.\begin{aligned}&t_N\approx(T/h)(1-e^{-1})=0.63T/h,\\&\max(r_N)\approx0.63R/h,\\&D(r_N)\approx R\sqrt{1-e^{-2}}/\sqrt{6h}=0.38\,R/\sqrt{h},\\&e_N=t_N-R_N,\end{aligned}\right\} \qquad (21.16)$$

因为 $y'''=-y''=y'=1-y$，因此，y''' 在 $1-Y(0)$ 与 $1-Y(1)$ 之間，即在 1.0～0.368 之間（参看表 22.1）。取 0.68 作为它的平均值，由(21.13)得到

$$T_n=(h^3/12)y'''=(0.68/12)h^3.$$

因此，最后得到

$$t_N=0.63\times0.68\times h^2/12=3.6\times10^{-4}.$$

---

❶　在这种情形，对应于 (21.11) 的是

$$D^2(r_n)=q^2D^2(r_{n-1})+(R^2/3)(1-hg/2)^{-2}, \quad D(r_0)=0.$$

由此就可以得到(21.14)。

設 $|R_N| \leqslant R=0.5\times 10^{-4}$，那么，$\max(r_N)=3.2\times 10^{-4}$，$D(r^N)=0.60\times 10^{-4}$。由此可以得出結論說，在点 $x=1.0$ 的誤差，由截断誤差引起的部分为小数第4位 $+3$ 或 $+4$ 左右，而由舍入誤差引起的部分不到小数第4位的1单位(参看表 22.1，梯形法則的 e 栏)。

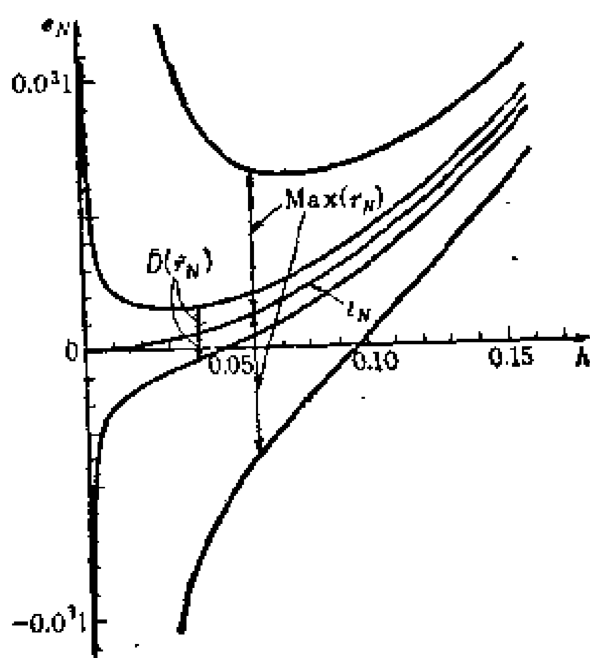


图 21.1

如果就目前的例子，令步长 h 取种种不同的值而画出 h 与誤差的关系，就得到图 21.1。一般說来，步长 h 縮小时，截断誤差的累积誤差 t_N 也跟着縮小，但由于計算的阶段数增大，因而由舍入引起的累积誤差 r_N 也跟着增大。图 21.1 是上述定性結論的量的表現之一。在本例中，最大的累积舍入誤差和累积截断誤差的和在 $h\approx 0.07$ 附近取得最小值。在这种极其簡单的情形这样就得到了最适步长的值，但在一般的情形，要决定最适步长的值是不容易的。

上面的处理法，是仿照 Crandall[11] 对于 Euler 方法的处理法而得到的。关于 Runge-Kutta 方法中的累积誤差的处理可以参看 Kopal[4]，关于 Milne 方法可以参看 Milne [14] 和其中引用的文献或柴垣 [17]。

§22 穩 定 性

在 §18 曾經說过，用改进的 Euler 方法进行多阶段的积分是不可靠的，图 22.1 就是这样的例子。它是解(21.15)所得的結果。由 $x=2.0$ 附近开始，图形显著地成为鋸齿形，且其振幅按指数規律增大。图中取 $h=0.1$，而把小数第5位作了四舍五入，但如进上去則較早就开始了振动，而舍去的話，开始得更早。同样取 $h=0.1$，如果小数第4位就四舍五入，振动开始得更早。取 $h=0.05$，則在 x 較大的地方才开始振动，振幅的增大也比較緩和。不論那一个，

在 x 的值較大的地方，和精确解 $1-e^{-x}$ 都是相差很远的。但是，这种状況不只限于 x 值較大的地方，在 x 值較小的地方已經开始了。

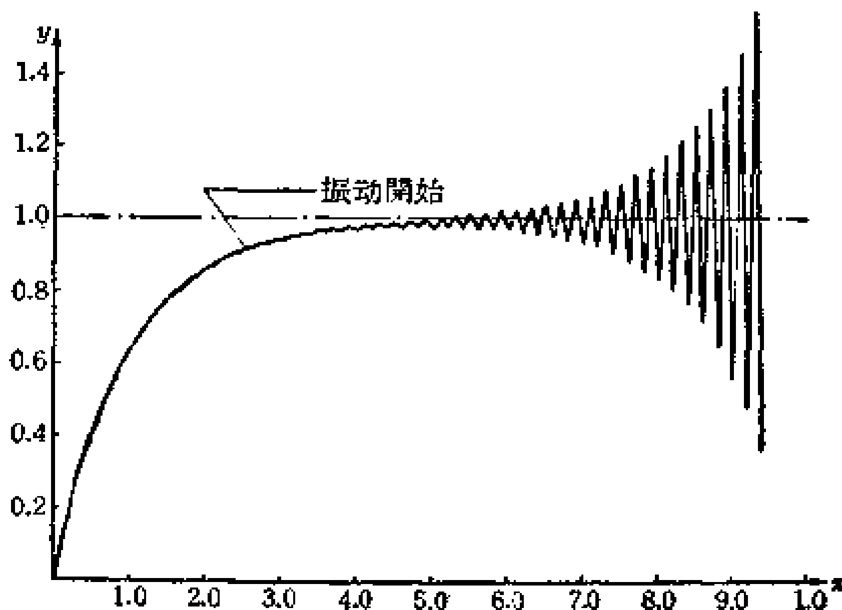


图　22.1

表 22.1　$y'=1-y,\ y(0)=0$.

	精确解 $1-e^{-x}$	Euler 方法 $y_{n+1}=y_n+hf_n$				改进的 Euler 方法 $y_{n+1}=y_{n-1}+2hf_n$				梯形法則 P: $y_{n+1}=y_{n-1}+2hf_n$ C: $y_{n+1}=y_n+(h/2)(f_n+f_{n+1})$			
x	y	Y	e	$\bar{Y}$	t	Y	e	$\bar{Y}$	t	Y	e	$\bar{Y}$	t
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.1	0.0952	0.1000	+48	0.1000	+48	0.0952	0	0.0952	0	0.0952	0	0.0952	0
0.2	0.1813	0.1900	+87	0.1900	+87	0.1810	−3	0.1810	−3	0.1814	+1	0.1814	+1
0.3	0.2592	0.2710	+118	0.2710	+118	0.2590	−2	0.2590	−2	0.2594	+0	0.2594	+0
0.4	0.3297	0.3439	+142	0.3439	+142	0.3292	−5	0.3292	−5	0.3299	+2	0.3299	+2
0.5	0.3935	0.4095	+160	0.4095	+160	0.3932	−3	0.3931	−3	0.3937	+2	0.3937	+2
0.6	0.4512	0.4686	+174	0.4686	+174	0.4506	−6	0.4505	−7	0.4515	+3	0.4515	+3
0.7	0.5034	0.5217	+183	0.5217	+183	0.5031	−3	0.5030	−4	0.5037	+3	0.5037	+3
0.8	0.5507	0.5695	+188	0.5695	+188	0.5500	−7	0.5499	−8	0.5509	+2	0.5510	+3
0.9	0.5934	0.6126	+192	0.6126	+192	0.5931	−3	0.5930	−4	0.5937	+3	0.5937	+3
1.0	0.6321	0.6513	+192	0.6513	+192	0.6314	−7	0.6313	−8	0.6324	+3	0.6324	+3

在小数第5位四舍五入。y：精确解，$\bar{Y}$：差分方程的解，Y：数值解，e：数値解的誤差，t：差分方程的解的誤差(截断誤差)。

这一点由到 $x=1.0$ 的数值解的誤差(表 22.1 e 栏) 就可以看出。表中的誤差是振动的,而且振幅有增大的趋势。

但是,在本例中, $f(x, y)=1-y$, 因此,代换微分方程的差分方程是❶

$$\overline{Y}_{n+1}=\overline{Y}_{n-1}+2h(1-\overline{Y}_n), \tag{22.1}$$

它的精确解是

$$\overline{Y}_n=1+A\lambda_1^n+B\lambda_2^n,$$

$$\lambda_1=\sqrt{1+h^2}-h, \quad \lambda_2=-(\sqrt{1+h^2}+h). \tag{22.2}$$

代入初始条件 $\overline{Y}_0=0, \overline{Y}_1=1-e^{-h}$ 而求出的值如表 22.1 $\overline{Y}$ 栏所示,它的誤差如 t 栏所示。与数值解的誤差 e 几乎完全相同。由此可見,誤差是截断誤差。(22.2) 的 λ_2 是大于 1 的負值,即使在开始时 $B=0$, 只要由于舍入等关系,这一項一旦出現,就会逐漸增大,而成为振动的原因。如果沒有含 λ_2 的項,数值解就接近于精确解。这种不必要的項出現的原因在于,原来的微分方程是一阶的,因而只需要一个初始条件,而代换它的差分方程却是二阶的。

由此可見,改进的 Euler 方法中含有不穩定的因素,因此在使用这方法时,必須注意 h 的大小,使在必要的范圍內誤差能小到所希望的程度。

对于(21.15)的例,不論用 Euler 方法还是用梯形法則,都得到差分方程的精确解:

$$\text{Euler 方法:} \quad \overline{Y}_n=1-(1-h)^n, \tag{22.3}$$

$$\text{梯形法則:} \quad \overline{Y}_n=1-(1-h/2)^n(1+h/2)^{-n}. \tag{22.4}$$

这些結果都記录在表 22.1 对应栏內 ($\overline{Y}$). 两者都和数值解几乎完全相同,由此可見,在本例中,数值解的誤差 e 主要是由截断誤

---

❶ 这里 $\overline{Y}_n$ 表示不作舍入时差分方程的解,以区别于作了舍入而求得的实际数值解 Y_n.

差引起的，舍入的影响几乎沒有表現出来。

由(22.3)，(22.4)可見，不論 Euler 方法还是梯形法則都不含有不稳定的因素。

§23 一阶方程組及高阶方程

只要是初始值問題，不論方程組也好，高阶方程也好，以上叙述的解一阶方程的方法都可以不变地使用。例如，用§19的方法(梯形法則，Milne 方法)解方程組

$$y'=f(x, y, z), \quad z'=g(x, y, z) \tag{23.1}$$

时計算的流动图如图 23.1(a) 所示。

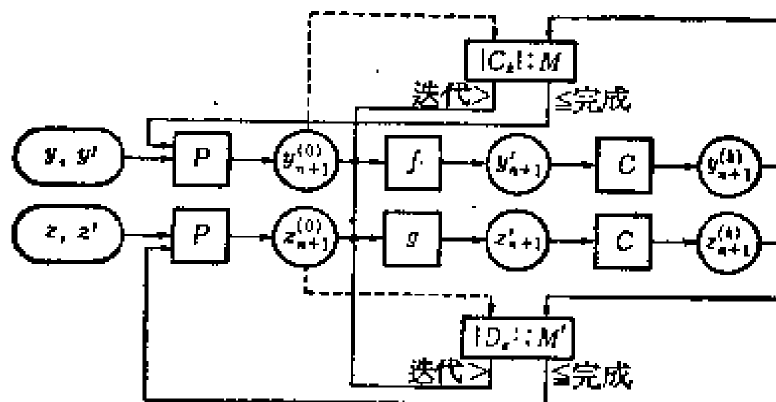


图 23.1(a) 解方程組 $y'=f(x, y, z)$，$z'=g(x, y, z)$ 的流动图，其中的記号参看图 19.1 关于 C_k，$D_k(=z_{n+1}^{(k+1)}-z_{n+1}^{(k)})$的諸条件同时满足时完了

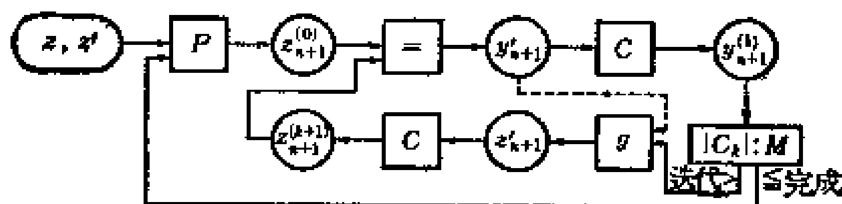


图 23.1(b) 解二阶微分方程 $y''=g(x, y, y')$ 的流动图．$y'=z$，$z'=g(x, y, z)$．判别在迭代的第2个循环($k=1$)以后进行

其次，高阶方程可以化成一阶方程組。例如

$$y''=g(x, y, y') \tag{23.2}$$

可以化为

$$y'=z, \quad z'=g(x, y, z). \tag{23.3}$$

只要在图 23.1(a) 中令 $f=z$ 就可以进行計算。但是,在高阶方程的情形,也可以使用图 23.1(b) 所示的簡化方法。在簡化法中預告算子只用于預告 $z(=y')$,而在計算 y 时不必使用。

例 1 $y''+xy'+y=0$, $y(0)=0$, $y'(0)=1$.

求解至小数第 6 位(精确至小数第 5 位)。写成(23.3)的形状:

$$y'=z, \quad z'=-xz-y, \quad y(0)=0, \quad z(0)=1.$$

展开成 Taylor 級数就得

$$y=x-\frac{x^3}{3}+\frac{x^5}{5\cdot 3}-\frac{x^7}{7\cdot 5\cdot 3}+\frac{x^9}{9\cdot 7\cdot 5\cdot 3}-\cdots,$$

$$z=1-x^2+\frac{x^4}{3}-\frac{x^6}{5\cdot 3}+\frac{x^8}{7\cdot 5\cdot 3}-\cdots.$$

令 $h=0.1$ 而求在 $x=0.1, 0.2, 0.3$ 处的值,以此作为最初的几个值而使用 Milne 方法,結果如表 23.1 所示。使用了图 23.1(a) 所示的方法。在本例中,迭代的次数最多是 3 次,h 好象取得过大了一些。用簡化法也試算了几个阶段,但迭代的次数完全相同,并未能使程序化簡。

二阶方程中常有不含 y' 項的,特别是在綫性方程的情形,常可以經过变換消去 y' 的項。例如,在方程

$$y''+P(x)y'+Q(x)y=R(x) \tag{23.4}$$

中,設 P, P', Q, R 都是連續的,那么,可以用下列两方法之一消去 y' 項。

1) 設 $y=Y\exp\left(-\int Pdx/2\right)$,将因变量 y 变为 Y,那么,方程(23.4)就化成

$$Y''+A(x)Y=0, \quad A(x)=Q-P'/2-P^2/4.$$

2) 設 $s=\int\exp\left(-\int Pdx\right)dx$,将自变量 x 变为 s,那么,方程(23.4)化成

$$\frac{d^2y}{ds^2}+B(s)y=0, \quad B(s)=Q(x)\exp\left(2\int Pdx\right).$$

表23.1 $y''+xy'+y=0$，$y(0)=0$，$y'(0)=1$ (Milne 方法) $h=0.1$

($y'=z$，$z'=-xz-y$，$y(0)=0$，$z(0)=1$)

	数值解								精确解
x	y	y'	$\frac{1}{3}\delta^2y'$	C_0	z	z'	$\frac{1}{3}\delta^2z'$	C_0	$e^{-\frac{1}{2}x^2}\int_0^x e^{\frac{t^2}{2}}dt$
0	0	1.000000			1.000000	0			0
0.1	0.099667	0.990033			0.990033	−0.198670			0.099667
0.2	0.197355	0.960529	−0.006512		0.960529	−0.389461	0.002626		0.197355
0.3	0.291160	0.912652	−0.006124		0.912652	−0.564956	5099		0.291160
0.4	0.379335	0.848265	−0.005503	+24	0.848265	−0.718641	7270	−30	0.379334
0.5	0.480345	0.769826	−0.004684	+21	0.769826	−0.845268	9023	−41	0.460344
0.6	0.532929	0.680240	−0.003716	+16	0.680240	−0.941074	10268	−52	0.532928
0.7	0.596131	0.582707	−0.002649	+12	0.582707	−1.004026	10954	−57	0.596128
0.8	0.649317	0.480543	−0.001543	+ 4	0.480543	−1.033751	11076	−58	0.649316
0.9	0.692194	0.377022	−0.000452	+ 3	0.377022	−1.031514	10654	−55	0.692192
1.0	0.724778	0.275220	−0.000570	− 5	0.275220	−0.999998	9793	−48	0.724778

其中，第 2 式右边的 x 应当由 $s=\int \exp\left(-\int P dx\right) dx$ 作为 s 的函数解出。

对于不含一阶导数的二阶微分方程，有可称之为二阶方程的 Milne 方法的下述解法。

$$
\left.\begin{aligned}
\text{P:}\quad & y_{n+1}-2y_{n-1}+y_{n-3}=4h^2(y''_{n-1}+\delta^2 y''_{n-1}/3) \\
& \qquad\qquad +h^6 y^{\text{VI}}(16/240), \\
\text{C:}\quad & y_{n+1}-2y_n+y_{n-1}=h^2(y''_n+\delta^2 y''_n/12) \\
& \qquad\qquad -h^6 y^{\text{VI}}/240.
\end{aligned}\right\} \qquad (23.5)
$$

截断誤差 T 由 $C_0/17$ 估計。使用这一公式所需要的最初的几个值与(17.6)相同，由下列的公式計算(柴垣 [17] p. 50)：

$$
\left.\begin{aligned}
y_1-y_0 &= hy'_0+\frac{h^2}{1440}(367f_0+540f_1-282f_2 \\
&\qquad +116f_3-21f_4)+\frac{107}{10080}h^7 y^{\text{VII}}, \\
y_2-y_0 &= 2hy'_0+\frac{h^2}{1440}(848f_0+2304f_1-480f_2 \\
&\qquad +256f_3-48f_4)+\frac{8}{315}h^7 y^{\text{VII}}, \\
y_3-y_0 &= 3hy'_0+\frac{h^2}{1440}(1323f_0+4212f_1+486f_2 \\
&\qquad +540f_3-81f_4)+\frac{9}{224}h^7 y^{\text{VII}}, \\
y_4-y_0 &= 4hy'_0+\frac{h^2}{1440}(1792f_0+6144f_1+1536f_2 \\
&\qquad +2048f_3+0)+\frac{16}{315}h^7 y^{\text{VII}}.
\end{aligned}\right\} \qquad (23.6)
$$

当然，也可以用 Taylor 展开等其他的方法。

上述計算方法与一阶方程情形的 Milne 方法完全相同。柴垣創立的方法和这方法几乎完全一样，只不过 P 的形状稍有不同而已。(参看柴垣[17] p. 112 以下。)

以下叙述一个利用綫性性质(解的迭加)的方法。

如果把微分运算用中心差分表示就得到(参看65頁)

$$\left.\begin{aligned}\frac{d}{dx}&\equiv D=\frac{1}{h}\Big(\mu\delta-\frac{\mu\delta^3}{3!}+\frac{2^2\mu\delta^5}{5!}-\frac{2^2\cdot 3^2\mu\delta^7}{7!}\\&\qquad+\frac{2^2\cdot 3^2\cdot 4^2\mu\delta^9}{9!}-\cdots\Big),\\\frac{d^2}{dx^2}&\equiv D^2=\frac{1}{h^2}\Big(\delta^2-\frac{2}{4!}\delta^4+\frac{2\cdot 2^2\delta^6}{6!}-\frac{2\cdot 2^2\cdot 3^2\delta^8}{8!}\\&\qquad+\frac{2\cdot 2^2\cdot 3^2\cdot 4^2\delta^{10}}{10!}-\cdots\Big).\end{aligned}\right\}\quad(23.7)$$

将这結果应用于二阶綫性微分方程

$$(a(x)D^2+b(x)D+c(x))y=f(x),\qquad(23.8)$$

而将三阶以上的差分移项到右边就得到

$$\left.\begin{aligned}L[y]&\equiv a(x)\delta^2y+hb(x)\mu\delta y+h^2c(x)y\\&=h^2f(x)+K[y],\\K[y]&\equiv\Big[hb(x)\Big(\frac{\mu\delta^3}{6}-\frac{\mu\delta^5}{30}+\cdots\Big)\\&\qquad+a(x)\Big(\frac{\delta^4}{12}-\frac{\delta^6}{90}+\cdots\Big)\Big]y.\end{aligned}\right\}\qquad(23.9)$$

作为第1近似,省略去 $K[y]$,而由

$$L[z^0]=h^2f(x)\qquad(23.10)$$

求 z^0. 由此第1近似 z^0 求得 $K[z^0]$,而由

$$L[z^1]=K[z^0]$$

求第1次的校正值。其次由 $L[z^2]=K[z^1]\cdots$ 求逐次的校正值。如果这些值的和 $z^0+z^1+z^2+\cdots$ 收斂的話,就得到

$$y=z^0+z^1+z^2+\cdots.$$

实际計算时,应该尽量将 h 取小,使得經过两次左右的校正以后就可以完成計算。

其次,因为需要中心差分,所以需要求解到求解范圍以外的某

些点(通常在两端点的外侧各取一点)。此外,作为初始值,給定 y_0 和 y_1 比之給定 y_0 和 y'_0 更为便利。由于綫性性质,这样給定初始值并不有损于一般性。

当这样給定初始值时,z^0, z^1, ⋯ 的初始值的取法如下:

$$z_0^0=y_0,\quad z_0^1=z_0^2=\cdots=0,$$

$$z_1^0=y_1,\quad z_1^1=z_1^2=\cdots=0.$$

关于这种处理方法,参看 Fox 和 Goodwin 的报告[1]以及 Milne [14], p. 93.

除了上述的方法以外,Runge-Kutta 的方法也可以用于求解方程組或高阶方程。在二阶方程的情形,为了求得更为精确的解,还有消去 y' 項而編制 y'' 的中心差分表来求解的方法。

§24 边界值問題

在求解方程組或高阶方程的問題中,常有在积分区間的两端点給出边界条件的情形。用解析方法求解时,即使是用需要初始值的 Laplace 变換求解,也可以先把初始值作为未定参数保留下来,等到作了反演之后,再由另一端点上的边界条件决定这些初始值,因而不发生什么問題[2]。但用数值方法求解时,必需首先知道必要的初始值 y_0, y'_0, y''_0, ⋯,或者在方程組的情形,y_0, z_0, w_0, ⋯ 等,否则就不能进行积分。

但是,如果是綫性方程,例如,形如

$$y''+a(x)y'+b(x)y=f(x)$$

的方程,就可以利用迭加性质处理如下。設 u, v 为当 $f=0$ 时对应的齐次方程的解,而 w 是上列方程的特解,那么,它的通解是

$$y=Au+Bv+w,$$

[1] Proc. Cambridge Phil. Soc., 45 (1949), pp. 373~388.

[2] 例如,近藤次郎:演算子法(培风館,昭和 31 年=1956 年),p. 75, 例 3.

其中 A, B 是积分常数。如果在点 $x=0$ 給定了一个条件,那么积分常数只剩下一个,因而上述通解实质上可以写成

$$y=Au+w$$

的形式。而且,$Au+w$ 应該看作是不論 A 的值如何都满足初始条件的解。也就是說,u 应該看作是除了条件 $u(0)=0$ 以外,再加上适当的初始条件而求出的对应齐次方程的数值解,w 是除了满足給定的初始条件以外,再加上适当的初始条件而求得的非齐次方程的数值解。于是,A 的値应当取得使綫性組合 $Au+w$ 满足另一端点上的条件,然后計算 $Au+w$ 就得到 y。

例 1 $$y''+xy'+y=2x, \quad y(0)=1, \quad y(1)=0. \tag{24.1}$$

表 24.1 (1) $y''+xy'+y=2x$, 在 $x=0$, $y=1$, 在 $x=1$, $y=0$.
(2) $y''+xy'+y=0$, 在 $x=0$, $y=0$, 在 $x=1$, $y=1$.

x	u	w	Au	y	y
0	0	1.000000	0	1.000000	0
0.1	0.099667	1.000000	−0.125908	0.874992	0.137514
0.2	0.197355	0.992061	−0.249316	0.742746	0.272297
0.3	0.291160	0.978436	−0.367818	0.610618	0.401723
0.4	0.379335	0.961498	−0.479204	0.482289	0.523381
0.5	0.460345	0.943652	−0.581547	0.362105	0.635153
0.6	0.532929	0.927232	−0.673242	0.253990	0.735300
0.7	0.596131	0.914418	−0.753084	0.161334	0.122502
0.8	0.649317	0.907158	−0.820273	0.086885	0.895884
0.9	0.692194	0.907112	−0.874439	0.032673	0.955043
1.0	0.724778	0.915602	−0.915602	0	1.000000

满足条件 $u(0)=0$ 的齐次方程的解 u 和方程 (24.1) 满足条件 $w(0)=1$ 的解 w 的綫性組合都能满足 (24.1) 的初始条件。满足 $u(0)=0$ 的解 u 如表 23.1 所示。其次,满足条件 $w(0)=1$ 的特解之一可以設 $w(0)=1$, $w(0.1)=1$ 而用上节(23.9)的方法求得如表 24.1 中 w 栏的样子(Milne[14],p. 96). 为了使它們的綫性組合 $y=Au+w$ 在点 $x=1$ 取值 0, 只要取 $A=-w(1)/u(1)$

$=-0.915602/0.724778=-1.263286$，这样决定 A 后，再計算 $v=Au-w$ (表24.1(1))。

例 2 应用上面所得的 u，齐次方程 $y''+xy'-y=0$ 滿足条件 $y(0)=0$，$y(1)=1.0$ 的解可以表示成 $y=Au$ 的形状，为此只要取 $A=1/u(1)$ 就可以求得(表24.1(2))。

在上述的綫性方程的情形，問題可以用解的迭加得到解决，但在非綫性的情形，这一方法不能应用。只能用試解的方法，即在另一端点上的条件不能滿足时，另外給定初始值，以求似乎能滿足上述边界条件的解。經过这样几次試解之后，由內插法或外插法可能得到适当的初始值。但有时也有不由某个特殊的初始值开始就得不到解的情形，也有完全得不到解的情形。这样，非綫性边界問題的求解是非常困难的。关于这类問題的解法虽有种种的尝試研究，但还沒有权威性的結果。毋宁說，化成变分問題而用直接(近似)方法求解可能还好一些。

如果把区間划分成网格而求解，在綫性方程的情形，就化成多元方程組的求解問題，这种方法与后面所述的椭圓型偏微分方程(Laplace 方程，Poisson 方程)的处理方法是相同的。

对于边界值問題，下节所述的近似解法在某些情形下也是很有效的。

§25 近 似 解 法[1]

以下叙述对求解边界值問題和特征值(eigen value)問題有用的几个近似解法。这些解法中，一般說来，Ritz 的方法是众所周知的，但比之以下叙述的方法，它的計算稍显繁杂，因此这里略去不談。

1) **Галеркин 方法** 这方法是 Б. Г. Галеркин 作为解薄壳和

[1] 关于本节討論的問題，参看本丛书《微分方程的近似解法》第4章。

梁的問題的方法在 1915 年提出来的，不但操作方法比較簡单，而且还可以用于解非綫性問題，因此得到广泛的应用。以下略去理論而只叙述它的方法❶。

假設要求綫性方程

$$L[y]=f(x) \tag{25.1}$$

在点 $x=a$, $x=b$ 满足給定边界条件的解。Галеркин 的方法就是选取满足上述边界条件的函数 $u_0(x)$，和在两端点 a, b 满足齐次边界条件的函数列 $u_1(x)$, $u_2(x)$, … 而假定所求的解 y 具有如下的形状：

$$y(x)=u_0(x)+a_1u_1(x)+a_2u_2(x)+\cdots+a_nu_n(x), \tag{25.2}$$

然后选取系数 $a_i(i=1, 2, \cdots, n)$，使 $y(x)$ 和精确解尽可能地接近。这条件可以代以如下的条件，即将 (25.2) 代入方程(25.1)时所得的殘差(residual)

$$R(x)=L\left[u_0(x)+\sum_{i=1}^{n}a_iu_i(x)\right]-f(x) \tag{25.3}$$

和上面选定的函数列 $u_i(x)$ $(i=1, \cdots, n)$ 正交，即

$$\int_a^b R(x)u_i(x)dx=0 \quad (i=1, 2, \cdots, n). \tag{25.4}$$

由此得到与待定系数 a_i 的个数 n 同数的 n 个代数方程（在目前考虑的綫性問題的情形，就是一次方程組）。由此可以定出 a_i，于是 y 也就跟着确定了。

例 1　对于边界条件 $y(0)=1$, $y(1)=0$, 可以取

$$u_0(x)=1-x, \quad u_i(x)=x^i(1-x) \quad (i=1, 2, \cdots, n),$$

$$y=(1-x)+a_1x(1-x)+a_2x^2(1-x)+\cdots+a_nx^n(1-x).$$

此时，計算(25.4)所需要的积分都具有

$$\int_0^1 x^m dx \quad (m=0, 1, \cdots)$$

---

❶ 关于 Галеркин 方法可以参看 Михлин 著《数学物理中的直接方法》一书。——譯者注

的形状，比之 Ritz 方法中出現的积分要簡单得多。

2) **山田方法**(矩量法)[1] 这方法是使殘差 $R(x)$ 和 x^i 正交而决定系数的方法，即令

$$\int_a^b R(x)x^i dx=0 \quad (i=0, 1, 2, \cdots, n-1). \qquad (25.5)$$

$u_0(x)$, $u_i(x)$ $(i=1, 2, \cdots, n)$ 滿足的边界条件与 Галеркин 方法相同，出現的积分的形状也是相同的。这是下述矩量定理的应用。所謂矩量定理就是說："在区間 (a, b) 上連續的函数 $f(x)$ 如果滿足形如

$$\int_a^b x^i f(x)dx=0 \quad (i=0, 1, 2, \cdots, n-1) \qquad (25.6)$$

的 n 个关系式，那么，$f(x)$ 在 (a, b) 上至少变号 n 次。也就是說，在 (a, b) 上有 n 个相异的根。在特例，当 $n\to\infty$ 时，$f(x)$ 在 (a, b) 上恒等于 0"。

山田方法比之 Галеркин 方法更为簡单，精确度也較好。这方法当然也可以用于非綫性方程，而且不只可以用于常微分方程，还可以用于偏微分方程、积分方程、差分方程，乃至函数在区間上的近似表示問題。特别是，在求特征值时是很便利的。

例 2 用山田方法解前节的例 (24.1)。設

$$y=(1-x)+a_1x(1-x)+a_2x^2(1-x)+a_3x^3(1-x),$$

計算 (25.5)，就得到关于 a_i 的如下方程組：

$$2a_1+a_2+a_3=-1,$$
$$65a_1+63a_2+62a_3=-50,$$
$$161a_1+189a_2+199a_3=-140.$$

由此求得，

$$y=(1-x)-0.2164x(1-x)-0.7703x^2(1-x)+0.2031x^3(1-x).$$

这样算出的 y 的数值，与用 Галеркин 方法算得的值无大差异，而在計算簡单一点上是比較优越的。如果設 y 为含有直到 a_4 的式子，就成为如下的样子：

---

[1] 山田彦児：九州大学流体力学研究所报告，3 卷 3 号(1947)。

$$y=(1-x)-0.216584x(1-x)-0.724468x^2(1-x)$$
$$-0.063174x^3(1-x)+0.094455x^4(1-x).$$

由此算出的值与表24.1所列的值比較，在小数第5位有等于 ±6 的誤差。

这些近似解法适用于求边界值問題的近似解。有时取 $n=2,3$ 左右就可以得到大致的估計，然后再用其他方法作精密計算。在非綫性的情形，虽然也可以使用，但其計算比之綫性的情形要繁杂。

象这样假定了函数形状的做法，和 Fourier 分解法很类似，因此，比之严格地划分成网格，化成差分方程而求解的方法尽管自由度(待定系数的个数)較少，但精确度很好，也是很自然的。这种想法也表現在給出了机翼理論中的具有奇异核的积分方程解法的 Multhopp[1] 和把这方法一般化了的近藤的"sample function"(抽样函数)[2] 的想法中。今后，这种想法也是値得推荐的。

---

[1] H. Multhopp: Luft. Forsch., 15 (1938), 153~169.

[2] K. Kondo: Journ. of the Faculty of Engineering, University of Tokyo, 24, 3 (March, 1955).

第6章　偏微分方程的数值解法

§26　偏微分方程的类型

直到現在，还沒有关于偏微分方程的一般的数值解法，只不过是对于每个特定的問題个别探求其解法而已。而且，过去主要只处理比較簡单的綫性方程，但是，最近由于大型电子計算机的使用，虽然范圍仍然有限，但非綫性的方程也逐漸可以求解了。

本章处理在热傳导乃至扩散問題中出現的抛物型方程，波动問題中出現的双曲型方程，以及在膜振动乃至圓柱体的扭曲問題中出現的 Laplace 方程，Poisson 方程等比較簡单的方程。至于解非綫性問題的例子，以及重調和方程，梁的弯曲振动方程等含有4阶导数的方程的解法，虽然也正在被探討中，但在这里不予涉及。

二阶綫性偏微分方程一般具有如下的形状[1]：

$$AU_{xx}+2BU_{xy}+CU_{yy}$$

$$=D(x,y)U_x+E(x,y)U_y+F(x,y)U+G(x,y). \quad (26.1)$$

U 右下方的字母表示关于这个字母的偏导数，A, B, C 也可能是常数，也可能是 x, y 的函数；方程的右边有时也可能是常数(包括0)。

按照系数 A, B, C 之間的关系，将方程(26.1)分类如下：

(i) $B^2-AC>0$ (双曲型)　此时方程(26.1)有2实特征綫

[1] 关于二阶綫性偏微分方程类型的一般討論請参看本丛书《偏微分方程》第5章。关于这里叙述的两变量的情形，还可以参看犬井鉄郎：应用偏微分方程論第2篇，第2章(岩波，1951)。

族。这种类型的微分方程的典型形状是下列的波动方程：

$$U_{tt}=a^2U_{xx}. \tag{26.2}$$

(ii) $B^2-AC=0$ (抛物型) 只有1族实特征綫，典型形状是下列的热傳导方程或扩散方程：

$$U_t=a^2U_{xx}. \tag{26.3}$$

(iii) $B^2-AC<0$ (椭圆型) 两族特征綫都不是实的，典型形状是

$$U_{xx}+U_{yy}=g(x,\ y) \quad (\text{Poisson 方程}), \tag{26.4}$$

$$U_{xx}+U_{yy}=0 \quad (\text{Laplace 方程}). \tag{26.5}$$

如果 A, B, C 是 x, y 的函数，方程的类型也可能因点 $(x,\ y)$ 的位置而异，不过，以下只就 (26.2) ~ (26.5) 的典型形状的简单情形加以討論。但是，复杂的情形在原則上仍是一样的。

問題的种类有，象热傳导方程、扩散方程或波动方程那样，給定了初始值和边界值，就可以依次进行积分的問題(对应于常微分方程的初始值問題)，以及象 Poisson 方程和 Laplace 方程那样，給定了区域边界上的条件，求区域内部的值的边界值問題。再者，例如热傳导問題，如果是周期的热傳导問題的話，也可以化成边界值問題加以处理。

§ 27 抛物型偏微分方程(前进型解法)

首先解最簡单的一維热傳导問題。方程具有如下的形状：

$$\frac{\partial\theta}{\partial t}=a^2\frac{\partial^2\theta}{\partial x^2}. \tag{27.1}$$

在热傳导問題中，θ 表示物体的溫度，a^2 是溫度傳播率，t 是时間而 x 是空間坐标。初始条件是

$$\text{当 } t=0 \text{ 时} \quad \theta=f(x), \tag{27.2}$$

边界条件一般以如下形状給出：

$$\begin{aligned}&\text{在 } x=0 \text{ 处} \quad \theta+\alpha_1\theta_x=\beta_1(t),\\ &\text{在 } x=1 \text{ 处} \quad \theta+\alpha_2\theta_x=\beta_2(t).\end{aligned} \tag{27.3}$$

以下用步长为 $\Delta x=h$, $\Delta t=k$ 的有限差分来代换偏微分。为此，假定 θ 在所考虑的 x, t 的区域内关于 x 有直到 6 阶的連續偏导数,而关于 t 有直到 3 阶的連續偏导数。作 Taylor 展开，就成为如下的形状:

$$\begin{aligned}&\theta(x+h,\ t)-2\theta(x,\ t)+\theta(x-h,\ t)\\ &\qquad =h^2\theta_{x^2}(x,\ t)+(h^4/12)\theta_{x^4}(x,\ t)\\ &\qquad\qquad +(h^6/360)\theta_{x^6}(s,\ t),\quad (x-h<s<x+h),\\ &\theta(x,\ t+k)-\theta(x,\ t)\\ &\qquad =k\theta_t(x,\ t)+(k^2/2)\theta_{t^2}(x,\ t)\\ &\qquad\qquad +(k^3/6)\theta_{t^3}(x,\ \tau),\quad (t<\tau<t+k).\end{aligned}$$

由 (27.1), $\theta_t=a^2\theta_{x^2}$, $\theta_{t^2}=a^4\theta_{x^4}\cdots$, 因此,如将上面的第一式乘以

$$r=ka^2/h^2, \tag{27.4}$$

而后由第二式减去所得的結果就得到

$$\begin{aligned}&\theta(x,\ t+k)-r\theta(x+h,\ t)-(1-2r)\theta(x,\ t)-r\theta(x-h,\ t)\\ &\quad =(h^4r/12)(6r-1)\theta_{x^4}(x,\ t)+(r^3h^6/6)\theta_{x^6}(x,\ \tau)\\ &\qquad -(rh^6/360)\theta_{x^6}(s,\ t).\end{aligned} \tag{27.5}$$

由(27.4)定义的 r 是表示 t 的步长与 x 的步长(的平方)之間的比例关系的参数。如果令 $r=1/6$, 那么,(27.5)右边 θ_{x^4} 的项也变成 0, 从而得到

$$\left.\begin{aligned}&\theta(x,\ t+k)=(1/6)\{\theta(x+h,\ t)+4\theta(x,\ t)\\ &\qquad\qquad\qquad +\theta(x-h,\ t)\}+T(x,\ t),\\ &T(x,\ t)=(h^6/1296)\theta_{x^6}(x,\ \tau)-(h^6/2160)\theta_{x^6}(s,\ t).\end{aligned}\right\} \tag{27.6}$$

也就是說,得到了截断誤差为 h^6 級的近似表示式。

由省去 $T(x,\ t)$ 一項的上列 $\theta(x,\ t+k)$ 的表示式,已知在时刻 t 及所有的点 x 的 θ 的值时，就可以得到 θ 在时刻 $t+k$ 的值，这

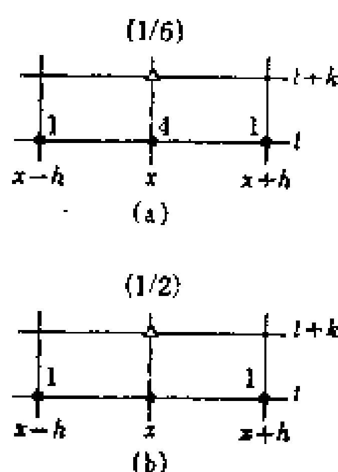


图 27.1

样，就可以依次进行积分。这是一种前进型的解法，它用网点表示的模型如图 27.1(a) 所示(图中，由在"·"处的值求在△处的值)。公式的形状也是簡单易用的。

此外，如令 $r=1/2$，就得到下列的虽然精确度不太好，但却更为簡单的公式：

$$\left.\begin{aligned}\theta(x, t+k) &= (1/2)\{\theta(x+h, t)\\ &\quad+\theta(x-h, t)\},\\ T &= (h^4/12)\theta_{x^4}.\end{aligned}\right\} \quad (27.7)$$

它用网点表示的模型如图 27.1(b) 所示，是极其簡单的。

例1 用公式(27.6)解

$$\frac{\partial\theta}{\partial t}=\frac{\partial^2\theta}{\partial x^2},$$

$$\theta(x, 0)=\sin\pi x,\ 0\leqslant x\leqslant 1;$$

$$\theta(0, t)=\theta(1, t)=0,\ t>0.$$

在表 27.1 中記載了开首几个阶段以及后来某些阶段的值。( ) 內的値是由精确解

$$\theta(x, t)=e^{-\pi^2 t}\sin\pi x$$

求得的，两者的值充分一致。

［注］ 象本例中的情形，由于沒有奇异性，因此，近似解和精确解极其相符。但如果把上面的初始条件换成 $\theta(x, 0)=1$，那么，在 $x=0$ 处当 $t=0$ 时出現不連續性。在这种有奇异点的情形，通常在奇异点的邻域內近似解和精确解不相符合。此时应当寻求代表这一奇异性的函数，然后就除去了奇异性的情形求数值解。此外，还应当說明，如把步长 h 取小，常常可以把奇异点的影响限制在局部范围內。

导出(27.6)的另一方法 上述的数学的处理方法詳見 Milne[14] p.121，高桥❶由物理的考虑导出了同一方法如下。

高桥的基本想法是，就扩散現象来說，在时刻 $t+k$ 的物理量，例如空气

❶ 高桥喜彦：气象集志，第2集，19卷 (1941)，pp. 321～327．又見：日高[16]下，pp. 274～276．

表 27.1 $\frac{\partial\theta}{\partial t}=\frac{\partial^2\theta}{\partial x^2}$, $\theta(x,0)=\sin\pi x$, $0\leqslant x\leqslant 1$;

$\theta(0,t)=\theta(1,t)=0$, $t>0$.

n \ m	0	1	2	3	4	5
0	0	0.30902	0.58778	0.80902	0.95106	1.00000
1	0	0.30398	0.57819	0.79582	0.93554	0.98369
2	0	0.29902	0.56876	0.78284	0.92028	0.96764
3	0	0.29414	0.55948	0.77007	0.90527	0.95185
4	0	0.28934	0.55036	0.75750	0.89050	0.93632
5	0	0.28462	0.54138	0.74514	0.87597	0.92104
6	0	0.27998	0.53254	0.73298	0.86168	0.90602
	(0)	(0.27998)	(0.53254)	(0.73299)	(0.86168)	(0.90602)
12	0	0.25366	0.48250	0.66409	0.78069	0.82086
	(0)	(0.25366)	(0.48249)	(0.66410)	(0.78070)	(0.82087)
18	0	0.22981	0.43714	0.60168	0.70732	0.74372
	(0)	(0.22982)	(0.43714)	(0.60169)	(0.70732)	(0.74372)
24	0	0.20821	0.39605	0.54512	0.64083	0.67380
	(0)	(0.20823)	(0.39606)	(0.54514)	(0.64085)	(0.67383)
30	0	0.18865	0.35882	0.49390	0.58059	0.61047
	(0)	(0.18866)	(0.35884)	(0.49390)	(0.58062)	(0.61050)

的温度 $\theta(x,t+k)$,是在时間 k 以前,即在时刻 t,在点 x 两傍邻近点的空气經过混和,因而其温度平均化的結果用式子写出就是

$$\theta(x,t+k)=\frac{1}{2h}\int_{x-h}^{x+h}\theta(x,t)dx. \tag{27.8}$$

把右边的积分用 Simpson 法則表出,就得到(27.6)。这里的 h 是,在时間 k 以内某一点的温度影响所及的从实用上来看的最大距离,k 愈大,或者 a^2(就空气的温度来說是渦动扩散率,就固体的热傳导来說是温度傳播率)愈大,h 也愈大。假如考虑分布在半无限空間中的空气,并設它的初始温度到处都是 0,在时刻 t,把它的表面 $x=0$ 驟然提高到某一温度,并設以后在这表面上一直保持同一温度,那么,在时間 k 以后的温度分布中,假如与表面 $x=0$ 距离不大于 h 的部分所含的热量,与由初始时刻 $t=0$ 到时刻 k 之間經过表面 $x=0$ 流入的全热量之比,例如說,是 95% 或 97%,那么就得,$h=2.4a\sqrt{k}$

或 $h=2.7a\sqrt{k}$. 这样就由物理的考虑决定了 h 和 k 的关系,如果把 h 取得很小,k 也非取小不可。这一比例系数在(27.6)中是 $\sqrt{6}=2.45$.

象这样的由物理的考虑而决定网格是很有意思的,而且这种想法也可以给出当 a 是 x 的函数时問題解法的某种启示,也可以用在后面叙述的稳定性的定性研究上。

一般說来,縮小步长 h 时,由(27.4)和它联系着的 k 与 h^2 成比例地縮小。因此,如果把 x 所取值的区間划分得很細时,t 的步长就变得很小,因而到达一定时間 t 的計算阶段数变得很多,这是不方便的。为了克服这种困难,可以考虑把 r 增大的方法。但由上述物理的想法也可以看到,增大 r 的方法是很不正确的。关于这个問題,在 § 29 中还要加以詳細的叙述。

§ 28 双曲型偏微分方程

考虑方程

$$\frac{\partial^2 u}{\partial t^2}=a^2\frac{\partial^2 u}{\partial x^2} \tag{28.1}$$

的求解問題。其中 a 表示波动的傳播速度。

这方程可以用下列的差分方程代换:

$$u_{m,n+1}-2u_{mn}+u_{m,n-1}=(a^2k^2/h^2)(u_{m+1,n}-2u_{mn}+u_{m-1,n}).$$

在特例,如果取 $ka=h$, 就得到

$$u_{m,n+1}=u_{m+1,n}+u_{m-1,n}-u_{m,n-1}. \tag{28.2}$$

这一差分方程具有与原来的微分方程(28.1)的精确解

$$u=f_1(x-at)+f_2(x+at)$$

形状相同的一般解。这一般解只要取 $ka=h$ 就与网格的大小无关。(但是,必須注意,如果取 h 大于扰动的波长,就有遺漏掉波长較短的扰动的可能。)网点的模型如图 28.1 所示。和上节所述类似,由到时刻 t 为止的值可以得到时刻 $t+k$ 的值,因而用前进型方法得到所

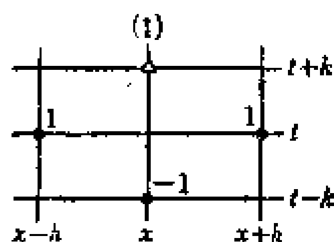


图 28.1

表28.1 $\frac{\partial^2 u}{\partial t^2}=a^2\frac{\partial^2 u}{\partial x^2}$. $u(x, 0)=0$, $\frac{\partial u(x, 0)}{\partial t}=0$, $0\leqslant x\leqslant 1$:

$$\left.\begin{array}{l} u(0, t)=\exp(-at/\alpha),\ \alpha=0.5, \\ u(1, t)=0, \end{array}\right\} t\geqslant 0$$

n \ m	0	1	2	3	4	5	6	7	8	9	10
0	1.0000	0	0	0	0	0	0	0	0	0	0
1	0.8187	1.0000	0								
2	0.6703	0.8187	1.0000	0							
3	0.5488	0.6703	0.8187	1.0000	0						
4	0.4493	0.5488	0.6703	0.8187	1.0000	0					
5	0.3679	0.4493	0.5488	0.6703	0.8187	1.0000	0				
6	0.3012	0.3679	0.4493	0.5488	0.6703	0.8187	1.0000	0			
7	0.2466	0.3012	0.3679	0.4493	0.5488	0.6703	0.8187	1.0000	0		
8	0.2019	0.2466	0.3012	0.3679	0.4493	0.5488	0.6703	0.8187	1.0000	0	
9	0.1653	0.2019	0.2466	0.3012	0.3679	0.4493	0.5488	0.6703	0.8187	1.0000	0
10	0.1353	0.1653	0.2019	0.2466	0.3012	0.3679	0.4493	0.5488	0.6703	0.8187	0
11	0.1108	0.1353	0.1653	0.2019	0.2466	0.3012	0.3679	0.4493	0.5488	−0.3297	0
12	0.0907	0.1108	0.1353	0.1653	0.2019	0.2466	0.3012	0.3679	−0.5507	−0.2699	0
13	0.0748	0.0907	0.1108	0.1353	0.1653	0.2019	0.2466	−0.6988	−0.4508	−0.2210	0
14	0.0608	0.0748	0.0907	0.1108	0.1353	0.1653	−0.7981	−0.5721	−0.3691	−0.1809	0
15	0.0498	0.0608	0.0748	0.0907	0.1108	−0.8647	−0.6534	−0.4684	−0.3022	−0.1481	0

求的解。

例　在初始条件：$t=0$ 时 $u=0$, $\partial u/\partial t=0$，和边界条件：在点 $x=0$, $u=\exp(-\alpha t/a)$, $\alpha=0.5$；在点 $x=1$, $u=0$ 下解微分方程

$$\partial^2 u/\partial t^2 = a^2\partial^2 u/\partial x^2.$$

在表 28.1 中記录了計算的某些阶段。

与前节所述抛物型方程的情形类似，(28.2) 也可以由物理的考虑导出❶。

§29　收斂性与穩定性

公式(27.6)的截断誤差比較小，計算也比較便利，但只要取定了 x 的步长 h，t 的步长 k 就是 $h^2/6a^2$，一般説来这是很小的值。为了补救这个缺陷，好象将 r 增大就可以了。以下首先令 $r=\dfrac{1}{2}$，利用(27.7)試解一个例題。

例 1　在初始条件：$t=0$ 时 $\theta=1$，和边界条件：在点 $x=0$, $\theta-\partial\theta/\partial x=0$，在点 $x=1$, $\partial\theta/\partial x=0$ 下解微分方程 $\dfrac{\partial\theta}{\partial t}=\dfrac{\partial^2\theta}{\partial x^2}$.

取 $h=1/4$ 而把在点 $x=0$ 的边界条件用差分式代换，就得到 $\theta_{-1,n}=\theta_{1n}-\dfrac{1}{2}\theta_{0n}$. 在表 29.1 中記录了一部分計算。

由物理意义来看，θ 应該是单調减小的，但上面的計算却得到振动的結果。如果把在点 $x=0$ 的边界条件代以 $\theta=0$，即使 $r=1/2$ 也不会得到如上的結果。这种現象的起因在于，$r=1/2$ 这个値正好是某个临界値，因而如果取 r 的値大于 $1/2$，不論在怎样的边界条件下，振动的振幅也会逐漸增大❷。这样，r 的値对于解的穩定性具有重要意义。

(27.5) 的右边去掉誤差項后的差分式是

$$\theta_{m,n+1}-r\theta_{m+1,n}-(1-2r)\,\theta_{mn}-r\theta_{m-1,n}=0. \qquad (29.1)$$

---

❶ 本問正作：气象集志，19卷19号，p. 351.

❷ 在这里，請回忆 §27 中叙述的网格 h 与 k 的大小的关系的物理意义。

表 29.1 $\frac{\partial\theta}{\partial t}=\frac{\partial^2\theta}{\partial x^2}$; $x=0$, $\theta-\frac{\partial\theta}{\partial x}=0$; $x=1$, $\frac{\partial\theta}{\partial x}=0$;

$t=0$, $\theta=1$. $\theta_{m,n+1}=\frac{1}{2}(\theta_{m-1,n}+\theta_{m+1,n})$,

$\theta_{-1n}=\theta_{1n}-\frac{1}{2}\theta_{0n}$, $h=\frac{1}{4}$.

n \ m	-1	0	1	2	3	4
0	0.5	1.0	1.0	1.0	1.0	1.0
1	0.625	0.750	1.0	1.0	1.0	1.0
2	0.46875	0.8125	0.875	1.0	1.0	1.0
3	0.57031	0.67188	0.90625	0.9375	1.0	1.0
4	0.43556	0.73828	0.80469	0.95312	0.96875	1.0
5	0.53564	0.62012	0.84570	0.88672	0.97656	0.96875
6	0.49808	0.69068	0.75342	0.91113	0.92774	0.97656
7	0.51053	0.58075	0.80090	0.84058	0.94384	0.92774
8	0.38280	0.65572	0.71066	0.87237	0.88416	0.94384
9	0.49068	0.54673	0.76404	0.79741	0.90810	0.88416
10	0.35839	0.62736	0.67207	0.83607	0.84078	0.90810
11	0.47410	0.51523	0.73172	0.75642	0.87208	0.84078
12	0.33437	0.60291	0.63582	0.80190	0.79860	0.87208
13	0.45986	0.48510	0.70240	0.71721	0.83699	0.79860
14	0.31059	0.58113	0.60116	0.76970	0.75790	0.83699
15	0.44748	0.45588	0.67542	0.67953	0.80334	0.75790

假定这里的 θ 可以分离成只含 x 的函数 $f(x)$ 和只含 t 的函数 $g(t)$ 的乘积：

$$\theta_{mn}=f_m\cdot g_n, \tag{29.2}$$

那么

$$(g_{n+1}-g_n)/g_n=r(f_{m+1}-2f_m+f_{m-1})/f_m. \tag{29.3}$$

左边的式中只含 n，右边的式中只含 m，因此，不論左边或右边实际上都不依赖于 m 及 n 而等于某一常数(設为 $-\varkappa$)，于是得到

$$g_{n+1}-(1-\varkappa)g_n=0, \tag{29.4}$$

$$f_{m+1}-(2-\varkappa/r)f_m+f_{m-1}=0. \tag{29.5}$$

由 (29.4) 得到如下形状的解：

$$g_n=a(1-\varkappa)^n\equiv a\lambda^n, \quad \lambda\equiv 1-\varkappa. \tag{29.6}$$

其中 a 由初始条件决定而 λ 由关于 x 的边界条件决定。令

$$2-\varkappa/r=2\cos\varphi, \tag{29.7}$$

那么 (29.5) 可以为 $\sin m\varphi$ 和 $\cos m\varphi$ 所满足。代入边界条件，就得到依赖于把 x 区間 L (在上例中是 1) 按长度 h 等分所得子区間个数 $M(=L/h)$ 的某个数那么多的 φ 值。这些 φ 值是特征值，而与此相对应的 $\sin m\varphi$ 和 $\cos m\varphi$ 是特征向量。

φ 的值例如可以成为如下的样子：

(a) 边界条件是：在点 $x=0.1$, $\theta=0$, 即 $\theta_{0n}=0$, $\theta_{Mn}=0$, 此时，

$$\varphi=j\pi/M \quad (j=1, 2, \cdots, M-1).$$

(b) 边界条件是：在点 $x=0$, $\theta=0$, 在点 $x=1$, $\partial\theta/\partial x=0$; 即 $\theta_{0n}=0$, $\theta_{M+1,n}=\theta_{M-1,n}$. 此时，

$$\varphi=(2j-1)\pi/2M \quad (j=1, 2, \cdots, M).$$

(c) 边界条件是：在点 $x=0$, $\theta-\partial\theta/\partial x=0$, 在点 $x=1$, $\partial\theta/\partial x=0$; 即 $\theta_{0n}-(\theta_{1n}-\theta_{-1n})/2h=0$, $\theta_{M+1,n}=\theta_{M-1,n}$. 此时，$\varphi$ 的值是满足下列关系式的 φ_j:

$$M\sin\varphi_j \operatorname{tg} M\varphi_j=1 \quad (j=1, 2, \cdots, M).$$

不論在那种情形，φ_j 都由边界条件决定。

由上所述可見，(29.1) 的解应該具有如下的形状：

$$\theta_{mn}=\sum_{j=1}^{M} a_j\{1-4r\sin^2(\varphi_j/2)\}^n v_j(m). \tag{29.8}$$

例如，在初始条件是：当 $t=0$ 时 $\theta=1$, 而边界条件是上述的 (a) 的情形，就得到 θ_{mn} 如下：

$$\theta_{mn}=\frac{2}{M}\sum_{j=1,3,\cdots}^{K}\left(\operatorname{ctg}\frac{j\pi}{2M}\right)\left(1-4r\sin^2\frac{j\pi}{2M}\right)^n \sin\frac{jm\pi}{M}, \tag{29.9}$$

其中，当 M 为偶数时 $K=M-1$, 而当 M 为奇数时，$K=M-2$. 与此相对应，原来的微分方程的解是

$$\theta(x, t)=\frac{4}{\pi}\sum_{j=1,3,\cdots}^{\infty}\frac{1}{j}e^{-j^2\pi^2 t}\sin j\pi x. \tag{29.10}$$

取 (29.9) 中依賴于时間 t 的項，得到

$$\begin{aligned}\lambda_j^n &= (1-4r\sin^2(j\pi/2M))^n \\ &= (1-4r\sin^2(j\pi h/2))^{t/rh^2}.\end{aligned} \tag{*}$$

当 r 固定而 $h\to0$ 时，只要 $r>0$，那么不論 r 取怎样的值，由(*)确定的 λ_j^n 都向 (29.10) 中的指数函数項收斂。因此，看来好象当 r 固定而 $h\to0$ 时差分方程的解向微分方程的解收斂，但是还不能严密保証差分方程的解对任意 r 都收斂。对于 $0<r<\frac{1}{2}$ 的情形，下面将証明它的收斂性，但对于 $r>\frac{1}{2}$ 的情形，則如后面将要指出的那样，除了收斂性問題以外，由于舍入誤差的产生和增大等影响而产生不稳定現象，这样的网格在实用上是完全沒有意义的。

关于上述的收斂性問題，在边界条件 (a)、初始条件 $\theta(x, 0)=f(x)$ (其中 $f(x)$ 是在区間 $(0, 1)$ 上除了在有限个点具有有限的跳跃外連續而且有界的函数) 下証明有如下的事实[1]。即，当 r 为滿足条件

$$0<r<\frac{1}{2} \tag{29.11}$$

的某个固定值时，对于区間 $(0, 1)$ 上的任意 x 和任意的 $t>0$，

$$\lim_{h\to0}\theta_{mn}=\theta(x, t).$$

这里 θ_{mn} 表示差分方程在和点 (x, t) 最近的网格点上的解。

已經証明，当 $f(x)$ 滿足更强的限制条件时，条件 (29.11) 的右边可以加上等号，即成为 $0<r\leqslant1/2$.

以上考虑了当关系于网格的比值 r 一定，而将网格无限細分时，差分方程的解 θ_{mn} 和微分方程的解 $\theta(x, t)$ 的差异，即相当于

---

[1] 例如，参看 F. B. Hildebrand: J. Math. Phys., 31. 1, pp. 35～41, 1952 或本节末文献 [3], p. 602.

解的誤差中由截断誤差引起的部分趋于 0 时，r 所必需滿足的条件。以下考虑数値解和假定未作舍入的 θ_{mn} 的差，即相当于由舍入誤差引起的解的誤差的部分。設在 $t=0$ 即在直綫 $n=0$ 上产生的誤差的分布为 $f(x)$，就得到和 (29.8) 形状完全相同的解 e_{mn}：

$$e_{mn}=\sum_{j=1}^{M} a_j \lambda_j^n v_j(m). \tag{29.12}$$

这里如果固定 h (因而也固定了 M) 而令 $n\to\infty$，若某个 λ_j 滿足 $\lambda_j<-1$，那么 e_{mn} 随着 n 的增大而振动，且其振幅无限增大，也就是說，誤差逐漸扩大地傳播。由差分得到的近似解在这种情况下称为是不稳定的。e_{mn} 为有界的必要条件是

$$\lambda_j=1-4r\sin^2(\varphi_j/2)\geqslant -1.$$

为此必須

$$r\leqslant (2\sin^2(\varphi_j/2))^{-1}. \tag{29.13}$$

当 r 滿足这一条件时，差分近似解是稳定的。对于边界条件 (a)，(b)，稳定性問題分別发生在 $j=M-1$ 和 $j=M$ 时，稳定的条件为

$$\left.\begin{aligned} &(\mathrm{a}): \quad r\leqslant \{2\sin^2((M-1)\pi/2M)\}^{-1},\\ &(\mathrm{b}): \quad r\leqslant (1+\cos(\pi/2M))^{-1}. \end{aligned}\right\} \tag{29.14}$$

如設 $M=10$，那么对于 (a) 就要求 $r\leqslant 0.513$ 而对于 (b) 就要求 $r\leqslant 0.503$。当 $h\to 0\ (M\to\infty)$ 时，稳定性条件就趋近于收斂性条件 (29.11)。此外，所有的項都不是振动的必要条件是 $\lambda_j>0$，为此，例如取由 (29.14) 等决定的 r 的上界的 1/2 左右，大体上可以滿足这一条件。在这个意义上，(27.6) 是足够稳定的。

[注] 把 (29.10) 的指数函数項改写而算出相当于 (*) 中的 λ_j 的 μ_j：

$$e^{-j^2\pi^2 t}=e^{-j^2\pi^2 nrh^2}=(e^{-j^2\pi^2h^2r})^n=(\mu_j)^n,$$

$$\therefore\quad \mu_j=1-(j\pi h)^2 r+(r^2/2)(j\pi h)^4+O(h^6).$$

另一方面，由 (*)，

$$\lambda_j=1-4r\sin^2(j\pi h/2)=1-(j\pi h)^2 r+(r/3\cdot 4)(j\pi h)^4+O(h^6).$$

由此看出，λ_j 和 μ_j 两者之間一般說来有 h^4 級的差异，只在 $r=\frac{1}{6}$ 时，有 h^6 級

的差异。与前述完全相符。

由上所述可以看出，稳定性依赖于代换微分方程以及边界条件的差分方程的性质，而在考虑誤差的傳播时就要考虑稳定性問題。关于稳定性和收斂性的关系参看本丛书«微分方程的近似解法»第6章。

关于收斂性与稳定性的詳細討論还可以参看下列的文献[1]，[2]，[3]和本书最后列举的 Crandall [11] pp. 380～385，Hildebrand ❶ 等参考书。关于边界条件的代换方法和稳定性可以参看文献[4]。

本节的参考文献

[1] R. Courant, K. Friedrichs and H. Lewy: Math. Ann., **100** (1928), 32～74.

[2] G. G. O'Brien, M. A. Hyman and S. Kaplan: Journ. of Math. and Phys., **29** (1950), 223～251.

[3] J. Todd: Comm. Pure and App. Math., **9** (1956), 597～612.

[4] P. H. Price and M. R. Slack: British Journ. of App. Phys., **3** (1952), 379～384.

§30 联立型解法

由前述物理的考虑也可以预料到，当网格的比值 r 过大时，解就会成为不稳定的。由定性方面的如下考虑也可以理解这种现象。考察前进型解法的过程，在图 30.1 中只要已知 AB 上的值，即使不知道与 P 同时刻的 CD 上

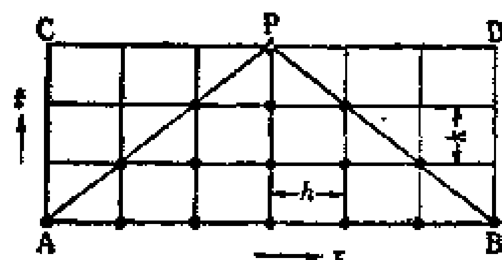


图 30.1

❶ Hildebrand [12], pp. 328～345 以及前頁脚注中的文献。

或者 AC, BD 上的值，在点 P 的值就确定了。也就是說，差分方程的解具有类似于双曲型方程的解的性质，而 AP, BP 可以看作是它的特征曲綫。但在实际上，由于微分方程是抛物型的，特征曲綫只有 $t=\text{const}$ 的直綫 CD, 因此，如果不知道 AC, BD 上的值，在点 P 的值应該是不能确定的。由此可以想到，在前进型的公式中，使用着与原方程的网点模型有所不同的网点模型。考虑到这一点，就得到下述的联立型解法。

这方法就是把直綫 $t=(n+1)k$ 上的值 $\theta_{m,n+1}(m=1, 2, \cdots, M-1)$ 作为未知数而列出关于它們的如下的联立方程組：

$$\left.\begin{aligned}&(\theta_{m,n+1}-\theta_{mn})/k=\{\alpha\delta_m^2\theta_{m,n+1}+(1-\alpha)\delta_m^2\theta_{mn}\}/h^2,\\ \text{其中}\quad &\delta_m^2\theta_{mn}=\theta_{m+1,n}-2\theta_{mn}+\theta_{m-1,n}.\end{aligned}\right\}\quad(30.1)$$

这里，α 是为了加权平均而作为权的常数，令 $\alpha=1$，就成为由 O'Brien 等人提出的逆三角型（▽ 型）解法[1]；令 $\alpha=\frac{1}{2}$，就成为 Crank 和 Nicolson 解非綫性方程时用过的 6 点型解法[2]。这些公式的形式如下：

▽ 型解法：$\theta_{m,n+1}-\theta_{mn}=r(\theta_{m+1,n+1}-2\theta_{m,n+1}+\theta_{m-1,n+1})$，

以及

$$\left.\begin{aligned}&\theta_{m+1,n+1}-(2+1/r)\theta_{m,n+1}+\theta_{m-1,n+1}=-\theta_{mn}/r,\\ &T=-\frac{rh^4}{2}\left(r+\frac{1}{6}\right)\theta_{x^4}.\end{aligned}\right\}\quad(30.2)$$

6 点型解法：

$$\left.\begin{aligned}\theta_{m,n+1}-\theta_{mn}&=\frac{r}{2}[(\theta_{m+1,n+1}-2\theta_{m,n+1}+\theta_{m-1,n+1})\\ &\quad+(\theta_{m+1,n}-2\theta_{m,n}+\theta_{m-1,n})],\\ T&=-\frac{rh^4}{12}\theta_{x^4}.\end{aligned}\right\}\quad(30.3)$$

---

[1] 参看前頁文献 [2]。

[2] J. Crank and P. Nicolson: Proc. Cambr. Phil. Soc., **43**, Part I (1947), pp. 50～67.

其中 T 表示截断誤差。

这些公式的网点模型如图 30.2 所示。如果把这些公式作 §29 中那样的处理，所得的 λ_j 具有如下的形状：

$$\left.\begin{aligned} &\nabla\text{ 型:}\\ &\qquad \lambda_j=(1+4r\sin^2(\varphi_j/2))^{-1},\\ &6\text{ 点型:}\\ &\qquad \lambda_j=(1-2r\sin^2(\varphi_j/2))(1+2r\sin^2(\varphi_j/2))^{-1}. \end{aligned}\right\}\quad(30.4)$$

由此可以看出，只要 $r>0$，常有 $|\lambda_j|<1$。但在后一公式的情形，如果 $r>(2\sin^2(\varphi_j/2))^{-1}$，那么就有 $\lambda_j<0$，因而解成为振动的，但其振幅并不增大。因此，不論 r 取怎样的值，解总是稳定的，如果只涉及稳定性問題，那么，t 的步长 k 不論取得怎样大都可以，但精确度不及 (27.6)。在某些問題中，有时非把 x 的步长 h 取小不可，而在必需計算到 t 的較大的值时，直接使用 (27.6) 进行計算需要非常多的阶段。在这一类情况下用 (27.6) 求出开始的几个阶段，而后用本节所述的联立型公式继續計算可能是适宜的。

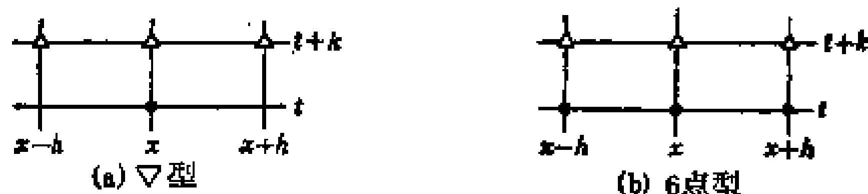


图 30.2

此外，应用 (30.2) 或 (30.3) 由直綫 $t=nk$ 上的值求直綫 $t=(n+1)k$ 上的值时，也可以用松弛法，在子区間的个数 M 不大时，也可以先求出逆矩陣而后使用松弛法。如果使用逆矩陣，那么它的程序与前进型无异。

[注 1] 在双曲型方程的情形，微分方程的特征曲綫与差分方程 (28.2) 的特征曲綫应当是一致的。如果 $k/h>1$，就得到差分方程一直到由微分方程給定的解的范圍以外的解，考虑到波动的性质，这是不合理的。已經証明

了，当 $k/h\leqslant 1$ 时，固定 k/h 而令 k, $h\to 0$，那么差分方程的解就向微分方程的解收敛。(141 頁文献 [1].)

［注 2］ 类似于 6 点法的想法，如果只把关于 t 的微分用差分代换而作成关于 x 的常微分方程，(27.1) 就可以代换成

$$\frac{d^2}{dx^2}\theta(x,\ t+k)-\frac{2}{a^2k}\theta(x,\ t+k)=-\frac{d^2}{dx^2}\theta(x,\ t)-\frac{2}{a^2k}\theta(x,\ t), \tag{30.5}$$

这是所謂 Hartree-Womersley 方法。他們用微分解析机解出了这方程。

上式还可以改写成

$$\frac{d^2\bar{\theta}}{dx^2}-\frac{2}{a^2k}\bar{\theta}=-\frac{2}{a^2k}\theta(x,\ t),\ \bar{\theta}=\frac{1}{2}\{\theta(x,\ t+k)+\theta(x,\ t)\}. \tag{30.6}$$

有一种以求原子反应堆的临界大小时所用的 multi-group 形式的想法为基础的計算，与 (30.6) 具有共同的操作方法。

不論那一种，都是知道了在时刻 t 的值，就可以求得在时刻 $t+k$ 的值。

也可以考虑只把关于 x 的微分用差分代换，而作为关于 t 的常微分方程求解的方法。

［注 3］ 簡单的前进型計算，用 UNIVAC-120 类型的机器就可以了。但是，如果是联立型，就需要如 ETL-Mark-II 等继电器式計算机 (relay calculator) 类型的記忆容量。有实际用 Mark-II 求解过的例子。

§31 2 維 算 子

以 x, y 为自变量的 2 維 Laplace 算子是

$$\nabla^2=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}. \tag{31.1}$$

如果把偏微分用关于 x, y 的差分代换而令步长 $\delta x=\delta y=h$，那么，∇^2 就成为(参看 (11.4) 或 (23.7))

$$\nabla^2=\frac{1}{h^2}\Big[\delta_x^2+\delta_y^2-\frac{2}{4!}(\delta_x^4+\delta_y^4)+\frac{2\cdot 2^2}{6!}(\delta_x^6+\delta_y^6)$$

$$-\frac{2\cdot 2^2\cdot 3^2}{8!}(\delta_x^8+\delta_y^8)+\cdots\Big]. \tag{31.2}$$

其中 δ 右下角的字母 x, y 表示关于对应字母的差分。用右边［ ］

內开首的两項表示 $\nabla^2\theta$ 就得到

$$\nabla^2\theta=\frac{1}{h^2}[\theta(x+h,y)+\theta(x-h,y)+\theta(x,y+h)+\theta(x,y-h)-4\theta(x,y)]+O(h^2). \tag{31.3}$$

它的网格的网点模型如图 31.1(a) 所示。(31.3) 右边 [ ] 内的算子用 H 表示,即❶

$$H=\begin{array}{|c|c|c|}\hline & 1 & \\ \hline 1 & -4 & 1 \\ \hline & 1 & \\ \hline\end{array} \tag{31.4}$$

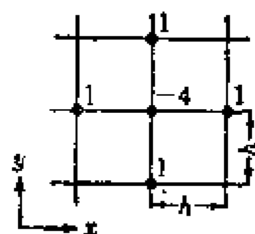


图 31.1(a) H 的网点模型

如将 x, y 軸旋轉 45° 而考慮图 31.1(b) 的网点模型,就得到

$$\nabla^2\theta=\frac{1}{2h^2}[\theta(x-h,y+h)+\theta(x+h,y+h)+\theta(x-h,y-h)+\theta(x+h,y-h)-4\theta(x,y)]+O(h^2). \tag{31.5}$$

右边 [ ] 内的算子用 $2X$ 表示,即❶

$$2X=\begin{array}{|c|c|c|}\hline 1 & & 1 \\ \hline & -4 & \\ \hline 1 & & 1 \\ \hline\end{array} \tag{31.6}$$

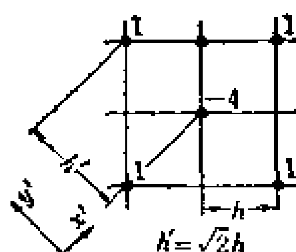


图 31.1(b) $2X$ 的网点模型

在某些情况下(特别在 Laplace 方程的情况下),把 H 和 X 結合起来比之单独使用 H, X 之一更能精确地得到算子 ∇^2 的近似表示。为此,首先把 H 和 X 中的 δ_x^2, δ_y^2 展开用 ∇^2 和 Q^4

❶ (31.4) 和 (31.6) 的右边有时叫做 stencil 或 lozenge. 还有,在《微分方程的近似解法》[19] §30 和 §31 中,代替 H 用記号 $\Diamond_h$,代替 $2X$ 用 $\Box_h$.

$\equiv \partial^4/\partial x^2\partial y^2$ 表示(参看 (11.10))：

$$H = h^2\nabla^2 + \frac{h^4}{2\cdot 3!}(\nabla^4 - 2Q^4) + \frac{h^6}{3\cdot 5!}(\nabla^6 - 3Q^4\nabla^2)$$

$$+ \frac{h^8}{4\cdot 7!}(\nabla^8 - 4Q^4\nabla^4 + 2Q^8) + \cdots, \tag{31.7}$$

$$X = h^2\nabla^2 + \frac{h^4}{2\cdot 3!}(\nabla^4 + 4Q^4) + \frac{h^6}{3\cdot 5!}(\nabla^6 + 12Q^4\nabla^2)$$

$$+ \frac{h^8}{4\cdot 7!}(\nabla^8 + 24Q^4\nabla^4 + 16Q^8) + \cdots. \tag{31.8}$$

为了使 h^2 项只含 ∇^2，h^4 只含 ∇^4，作算子

$$K = 4H + 2X, \tag{31.9}$$

就得到

$$\frac{1}{6}K = h^2\nabla^2 + \frac{h^4}{2\cdot 3!}\ \nabla^4 + \frac{h^6}{3\cdot 5!}(\nabla^6 + 2Q^4\nabla^2)$$

$$+ \frac{h^8}{4\cdot 7!}\left(\nabla^8 + \frac{16}{3}Q^4\nabla^4 + \frac{20}{3}Q^8\right) + \cdots. \tag{31.10}$$

在用于 Laplace 方程 $\nabla^2\theta = 0$ 时，可以用 $K/6h^2$ 近似表示 ∇^2 到 h^6 级的精确度(用 H，X 表示时精确度是 h^2 级)。K 的网点模型如图 31.2 所示，它的 stencil 如下：

$$K = \begin{array}{|c|c|c|} \hline 1 & 4 & 1 \\ \hline 4 & -20 & 4 \\ \hline 1 & 4 & 1 \\ \hline \end{array} \tag{31.11}$$

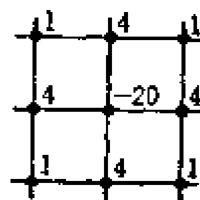


图 31.2　K 的网点模型

其次，导入近似表示 $Q^4 \equiv \partial^4/\partial x^2\partial y^2$ 的算子

$$N^2 \equiv \delta_x^2\delta_y^2 = \begin{array}{|c|c|c|} \hline 1 & -2 & 1 \\ \hline -2 & 4 & -2 \\ \hline 1 & -2 & 1 \\ \hline \end{array} \tag{31.12}$$

用 ∇^2, Q^4 展开就得到

$$N^2 = h^4 Q^4 + \frac{h^6}{2\cdot 3!} Q^4 \nabla^4 + \frac{h^8}{3\cdot 5!} Q^4 \left(\nabla^4 + \frac{1}{2} Q^4\right) + \cdots, \tag{31.13}$$

N^2 与 H, X 之間有如下的关系:

$$N^2 = 2(H - X). \tag{31.14}$$

由 (31.9), (31.14), 用 K, N^2 表示 $H(=\delta_x^2+\delta_y^2)$, 就得到 $\delta_x^2+\delta_y^2=(K-N^2)/6$, 以及 $\delta_x^4+\delta_y^4=(K-N^2)^2/36-2N^2$, 因此, (31.2) 的 ∇^2 可以用 K 和 N^2 表示如下:

$$\nabla^2 = \frac{1}{6h^2}\left[K - \frac{K^2}{72} + \left(\frac{K^3}{3240} - \frac{KN^2}{180}\right) - \left(\frac{K^4}{120960} - \frac{K^2N^2}{3780} + \frac{N^4}{504}\right) + \cdots\right], \tag{31.15}$$

而

$$Q^4 = \frac{1}{h^4}\delta_x^2\delta_y^2\left\{1 - \frac{1}{12}(\delta_x^2+\delta_y^2) + \frac{1}{90}(\delta_x^4+\delta_y^4) + \frac{1}{144}\delta_x^2\delta_y^2 - \cdots\right\}$$
$$= (N^2/h^4)\{1 - K/72 + K^2/540 - N^2/120 - \cdots\}. \tag{31.16}$$

二維的热傳导方程和扩散方程, 或者波动方程可以用上述的 H 和 K 表示。例如,热傳导方程

$$\frac{\partial\theta}{\partial t} = a^2\nabla^2\theta \tag{31.17}$$

利用上述的算子可以改写成

$$\theta(x, y, t+k) = k^2 a^2 \nabla^2\theta + \theta(x, y, t) \approx a^2 \frac{k}{h^2} H\theta + \theta.$$

在此,如果令 $r \equiv a^2k/h^2 = 1/6$, 就得到

$$\theta(x, y, t+k) = \left(\frac{1}{6}H + 1\right)\theta = \frac{1}{6}\begin{array}{|c|c|c|}\hline & 1 & \\ \hline 1 & 2 & 1 \\ \hline & 1 & \\ \hline\end{array}\,\theta(x, y, t), \tag{31.18}$$

相当于一維情形的 (27.6)。如果用算子 K，就得到

$$\theta(x,\ y,\ t+k)=\left(\frac{1}{36}K+1\right)\theta=\frac{1}{36}\begin{array}{|c|c|c|}\hline 1 & 4 & 1\\ \hline 4 & 16 & 4\\ \hline 1 & 4 & 1\\ \hline\end{array}\theta(x,y,t).$$

(31.19)

Yowell 指出，上式最后的 stencil 可以考虑作如下的乘积形式(参看 Milne [14] p. 138)：

$$\underbrace{\left\{\frac{1}{6}\begin{array}{|c|c|c|}\hline 1 & 4 & 1\\ \hline\end{array}\right\}\underbrace{\left\{\frac{1}{6}\begin{array}{|c|}\hline 1\\ \hline 4\\ \hline 1\\ \hline\end{array}\right\}\theta(x,y,t)}_{\text{第1段}}}_{\text{第2段}}$$

这一事实在使用穿孔卡或紙帶的計算机中，将二維方程改写成直列式时是有用的。

(31.19) 的截断誤差是

$$T=\left(\frac{K^3}{699840}-\frac{KN^2}{6480}\right)\theta(x,\ y,\ t)+O(h^8).\qquad(31.20)$$

如果近似地认为，$K/36\approx(1/6)h^2\nabla^2=k\partial/\partial t\approx\Delta_t$ (Δ_t 表示关于 t 的差分)，那么

$$T\approx\left(\frac{\Delta_t^3}{15}-\frac{\Delta_t N^2}{180}\right)\theta(x,\ y,\ t).\qquad(31.21)$$

由此可以近似估計計算結果的誤差。

在解前进型的問題时用 K 也是可以的，但在解后述的 Laplace 或 Poisson 方程时用 H 或者 X 是比較簡单的，而在以后的驗算中則应当用 K。

此外，二維的波动方程也可以用 K 或 N^2 改写成差分方程，但由于应用的机会不多，这里略去和一維波动方程的 (28.2) 相对应的二維差分方程。

§ 32 关于椭圆型方程的解法

如前所述，热傳导(抛物型)方程或波动(双曲型)方程可以用前进型方法求解，具有和常微分方程的初始值問題(或过渡現象問題)类似的性质。与此相对应，椭圆型偏微分方程和常微分方程的边界值問題有类似的性质，而其解法和前两种类型的偏微分方程完全不同。

考虑求 2 維 Laplace 方程取給定边界值的解的問題。設在如图 32.1 所示的边界內部，函数 u 滿足 Laplace 方程

$$\nabla^2 u=0, \tag{32.1}$$

而在边界 Γ 上取某些确定的值 $b_i\,(i=1, 2, \cdots)$. 将图 32.1 中的区域分成如图所示的网格，并且如图中所示那样确定网点的編号和 u 在 Γ 上所取值的記号；代替(32.1) 的 Laplace 算子使用 H 而作方程

$$H(u_i)=0 \quad (i=1, 2, \cdots, 10), \tag{32.2}$$

就得到方程組(32.3)：

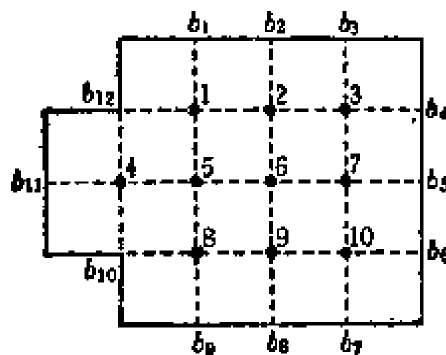


图 32.1

$$
\left.\begin{array}{l}
4u_1-u_2-u_5=b_1+b_{12},\\
-u_1+4u_2-u_3-u_6=b_2,\\
-u_2+4u_3-u_7=b_3+b_4,\\
4u_4-u_5=b_{10}+b_{11}+b_{12},\\
-u_1-u_4+4u_5-u_6-u_8=0,\\
-u_2-u_5+4u_6-u_7-u_9=0,\\
-u_3-u_6+4u_7-u_{10}=b_5,\\
-u_5+4u_8-u_9=b_9+b_{10},\\
-u_6-u_8+4u_9-u_{10}=b_8,\\
-u_7-u_9+4u_{10}=b_6+b_7.
\end{array}\right\}
\tag{32.3}
$$

在此令

$$
H=\begin{pmatrix}
4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0\\
-1 & 4 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0\\
0 & -1 & 4 & 0 & 0 & 0 & -1 & 0 & 0 & 0\\
0 & 0 & 0 & 4 & -1 & 0 & 0 & 0 & 0 & 0\\
-1 & 0 & 0 & -1 & 4 & -1 & 0 & -1 & 0 & 0\\
0 & -1 & 0 & 0 & -1 & 4 & -1 & 0 & -1 & 0\\
0 & 0 & -1 & 0 & 0 & -1 & 4 & 0 & 0 & -1\\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 4 & -1 & 0\\
0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 & -1\\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4
\end{pmatrix}
\tag{32.4}
$$

就得

$$
\boldsymbol{Hu}=\boldsymbol{b}. \tag{32.5}
$$

其中 $\boldsymbol{u}$ 是以 $u_i(i=1, 2, \cdots, 10)$ 为分量的列向量，而 $\boldsymbol{b}$ 是以 (32.3) 的右边为分量的列向量。

在 Poisson 方程

$$\nabla^2 u = \rho(x, y) \tag{32.6}$$

的情形，(32.5) 右边的 $\boldsymbol{b}$ 中包含 $\rho(x, y)$ 在各个网点上的值。

这样，問題就归結为解关于 u_i 的联立方程組。如果問題是关于左边相同，給定种种 $\boldsymbol{b}$ 的值而求解，那么求出 $\boldsymbol{H}$ 的逆矩陣是有益的。但一般說来，用手工計算时常用松弛法，而在內部网点数 n 很大需用机械計算时，較适宜的是使用后述的几个逐次逼近法。为此，需要知道 $\boldsymbol{H}$ 的性质(例如，它的特征值等)。下面列举矩陣 $\boldsymbol{H}$ 的某些已知性质：

1. 矩陣 $\boldsymbol{H}$ 不依賴于边界值 b_i，而只依賴于区域 D 內以及边界 Γ 上网点的分布状况。

2. 矩陣 $\boldsymbol{H}$ 是对称的。

3. 矩陣 $\boldsymbol{H}$ 的特征值关于 $\lambda=4$ 是对称的，即，如果 λ_i 是它的一个特征值，那么，$8-\lambda_i$ 也是它的一个特征值。由此推出，如果內部网点的个数是奇数，那么 4 是一个特征值。而且，一切特征值 λ 都在 $0<\lambda<8$ 的范圍內。

4. $|\boldsymbol{H}| \neq 0$，也就是說，$\boldsymbol{H}$ 是正規的。

这些事实在用机械計算进行迭代而求解时，可以加快解的收斂性。如果代替 H 而使用 K 也可以得到和(32.5)形式相同的式子。此时，关于代替 $\boldsymbol{H}$ 而出現的矩陣 $\boldsymbol{K}$，也有与 $\boldsymbol{H}$ 类似的种种已知性质，但 K 最好不直接用于提高精确度，而用于驗算。为了提高精确度，可以把步长縮小而仍用算子 H 是适宜的。

以下叙述用于机械計算的迭代法，其原理与以前在 § 3 中所述的一般联立方程的解法是相同的。

考虑 Poisson 方程在网点 (i, j) 改写成差分方程的形状：

$$4u_{ij} - u_{i+1,j} - u_{i,j+1} - u_{i-1,j} - u_{i,j-1} = h^2\rho_{ij}. \tag{32.7}$$

将此式改写成便于迭代的形式就得

$$u_{ij} = \frac{1}{4}\{u_{i+1,j} + u_{i,j+1} + u_{i-1,j} + u_{i,j-1} + h^2\rho_{ij}\}. \tag{32.8}$$

这样的式子共有和网点个数相同的个数，而且 u 在每个网点的值可用它在邻近4个网点的值表示出来。因此，作为联立方程組，虽然方程的个数很多，但每一个方程却很簡单，因此可以說，用迭代法求解是适宜于机械計算的。

作为上述迭代法之一，用第 k 近似值代右边的一切值而求第 $k+1$ 近似值，就是所謂 Richardson 迭代法。即

$$u_{ij}^{(k+1)}=\frac{1}{4}\{u_{i+1,j}^{(k)}+u_{i,j+1}^{(k)}+u_{i-1,j}^{(k)}+u_{i,j-1}^{(k)}+h^2\rho_{ij}\}. \tag{32.9}$$

这一方法是 Richardson 在1910年計算水坝的应力时所用，方法虽然簡单，但有收斂很慢的缺点。

这一方法是把右边的值全部一次代以新的近似值的方法，而 H. Liebmann❶ 則提倡用在計算过程中已經得到的最新的值的方法。在这方法中，首先从例如和边界最近的内側的点那样，右边的未知数尽可能少的点开始，在和这些点最邻近的点，立刻就使用刚才得到的值。也就是說，如果已經得到了 $u_{i+1,j}$，$u_{i,j+1}$ 的第 $(k+1)$ 近似值，那么，u_{ij} 的第 $(k+1)$ 近似值 $u_{ij}^{(k+1)}$ 由下式給出：

$$u_{ij}^{(k+1)}=\frac{1}{4}\{u_{i+1,j}^{(k+1)}+u_{i,j+1}^{(k+1)}+u_{i-1,j}^{(k)}+u_{i,j-1}^{(k)}+h^2\rho_{ij}\}. \tag{32.10}$$

例如，如果在图32.1中，按网点編号的次序进行迭代，那么計算在网点2的值时，就要使用已經得到的在点1的值。这方法比之 Richardson 方法一般說来收斂較快。乍看起来虽然似乎复杂，但由于可以把計算机中的存储逐步代以新得到的值，反而是便于机械計算的。

这一方法作为方程組的解法来說，就是 Gauss-Seidel 方法 (14頁)。

不过，在网点的个数很大时使用这方法收斂仍旧是很慢的。例如，关于两边为 ph，qh 的矩形区域 (h 表示步长，即把矩形的两边分别 p 等分，q 等分了的情形)，設为了得到精确度 r 所需要的迭代次数为 N，那么，据說有

$$N=\frac{r}{2}\Big/\left[-\log\left\{\frac{1}{2}\left(\cos\frac{\pi}{p}+\cos\frac{\pi}{q}\right)\right\}\right]❷. \tag{32.11}$$

认为 p，q 都很大而由 (32.11) 求出 N 的近似值就得到

$$N\approx p^2q^2(p^2+q^2)^{-1}r(\ln 10)/\pi^2,$$

❶ H. Liebmann: Sitz. bayer. Akad. Wiss. Math.-Phys. Klasse 3 (1918), pp. 385～416. 又見，日高 [16] 下 pp. 86～100.

❷ S. P. Frankel: Math. Tables and Other Aids to Comput., 4 (1950), pp. 65～75.

因此，如果把 h 縮小，N 就与 h^{-2} 成比例地增大，而且內部网点的个数也和 h^{-2} 成比例地增加，整个說来，計算量有 h^{-4} 級的增长，因此，卽使使用机械計算也还是很可观的。为了解决这問題，正在考虑加速收斂性的方法。

作为这些方法之一，以下簡单叙述加快 Liebmann 方法。(詳細的說明可以参看概括介紹 n 很大的椭圓型方程解法的 E. L. Wachspress 的报告❶。) 这是前述 Liebmann 方法的改进，叫做 accelerated Liebmann 方法或 extrapolated Liebmann 方法。通常的 Liebmann 方法就是直接使用 $u_{ij}^{(k+1)}$ 和 $u_{ij}^{(k)}$ 的差，卽殘差 (residual)

$$R_{ij}^{(k)}=\frac{1}{4}\{u_{i+1,j}^{(k+1)}+u_{i,j+1}^{(k+1)}+u_{i-1,j}^{(k)}+u_{i,j-1}^{(k)}+h^2\rho_{ij}\}-u_{ij}^{(k)} \tag{32.12}$$

而令

$$u_{ij}^{(k+1)}=u_{ij}^{(k)}+R_{ij}^{(k)}. \tag{32.13}$$

加快 Liebmann 方法則是利用某个常数 α 校正上式，而令

$$u_{ij}^{(k+1)}=u_{ij}^{(k)}+\alpha R_{ij}^{(k)}. \tag{32.14}$$

选取

$$\alpha=2(1+\sqrt{1-\lambda_n^2})^{-1} \tag{32.15}$$

时收斂得最快，其中 λ_n 是把 (32.8) 表示成

$$u=Lu+\frac{1}{4}h^2\rho \tag{32.16}$$

的形状时矩陣 L 的最大特征值。α 的值因 L 而异，也就是因問題而异，但一般說来，$1\leqslant\alpha\leqslant2$，特别是，令 $\alpha=1$，就得到通常的 Liebmann 方法。(α 和 2 越接近，越能发揮加快 Liebmann 方法的优越性。) 为了确定适当的 α，成問題的是求 λ_n，但并不一定要求 λ_n 的精确值，只要求得它的近似值就够用了。

这一方法在計算原子反应堆的临界量，以及作气象的数值預报等需要很多网点的情形是极其有用的。

最后，叙述一下边界凹凸的影响。

为了說明这一問題，首先用如图 32.1 所示那样較粗的网格进行計算，而后在其边界附近，如图 32.2 所示把网格进一步細分。

---

❶ Iterative methods for solving elliptic-type differential equations with application to two-space-dimension multi-group analysis, KAPL-1333 (Knolls Atomic Power Laboratory), 1955.

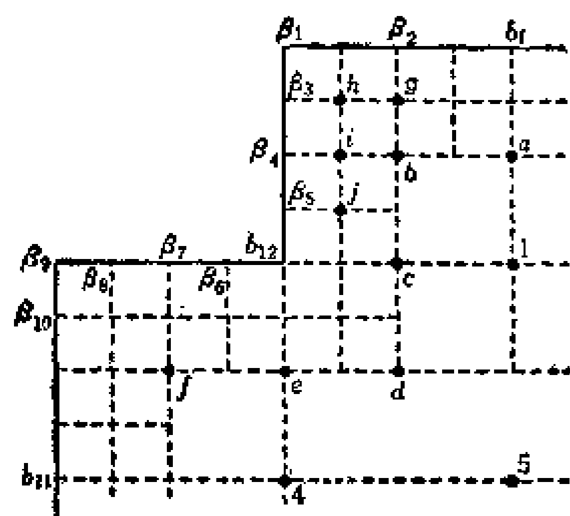


图 32.2

求此时新出现的网点 $a, b, \cdots$ 上的值的程序如下：

$$u_a=\frac{1}{2}(u_1+b_1),$$

$$u_b=\frac{1}{4}(u_1+b_1+\beta_1+b_{12}),$$

$$u_c=\frac{1}{2}(u_1+b_{12}),\ \cdots.$$

以这样求得的值为基础，再用松弛法或其他方法計算 u_a, u_b, …等。此时 u_1, u_2, …等也发生变化。将这些值全部算出以后，再用同上的程序計算 u_g, u_h, u_i, …，最后以这些值为基础用松弛法等求在各网点上的值。

这样进行計算时，一般說来，点离边界越近，这些点上的值对于內部的影响越小，因此，u_1 以下的值的变化不大。

上述的程序不仅可用于边界附近，在用較粗的网格进行計算后，为了提高精确度，对于內部的网点也可以适用。

上面所說的是网点正好在边界上的情形，但在曲綫边界的情形，网点多半不在边界上，此时，对于在边界附近的点的 Laplace 算子的近似表示式也有种种倡議，但一般說来計算既复杂而且精确度也不好。在这种情形，在边界附近把网格进一步細分的方法还是比較好的。

关于网格的形状，也有試图用正3角形网或正6角形网的❶。还有，对于不是 (x, y) 平面上而是軸对称的問題也有用圓柱坐标 (r, z) 的网格的尝試❷。

---

❶ 例如本丛书《微分方程的近似解法》§31 或 Kunz [8] 第12章。

❷ G. Shortley 等：Journ. of App. Phys., 18 (1947), pp. 116～129.

§33 Poisson 型方程用松弛法的解法

以下叙述更适用于手工計算的、在很大程度上依赖于計算工作者的判断与熟练程度的松弛法 (relaxation method) 的例子。

例 1 試解如下在正方形区域内的 Poisson 型方程:

$$\left.\begin{array}{ll}\nabla^2 u=-2 & (-1\leqslant x\leqslant 1,\ -1\leqslant y\leqslant 1),\\ \text{在边界}\,(x=\pm 1\text{ 和 }y=\pm 1)\text{ 上} & u=0,\end{array}\right\}\tag{33.1}$$

这是在彈性力学中正方柱的扭曲中出現的問題。

將上式改写成形如 (32.5) 的差分方程就得到

$$\boldsymbol{H}\boldsymbol{u}=-2h^2,\tag{33.2}$$

其中 h 表示步长。以下用松弛法来解 (33.2)。

首先叙述一下大体的程序。

程序 1　由边界的形状，边界值等尽可能地利用对称性減少必需的网点个数。

程序 2　首先用較粗的网格求出在各网点的最初的估計值。这些估計值可以由边界值按簡单的比例关系求出。

程序 2′　如果較粗网格的网点数約为 3～4 个，那么用通常的消去法解 $\boldsymbol{H}\boldsymbol{u}=\boldsymbol{b}$ (在本例中即 (33.2))，就得到在較粗网格上的精确值。此时，程序 2～4 一下就完成了。

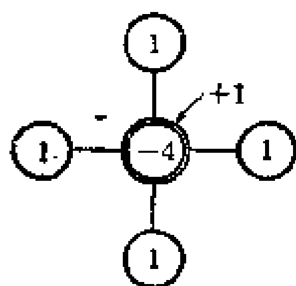
程序 3　在网点 (i, j) 的最初的估計值，当然不滿足 (33.2)。設在中心点的殘差为 R_{ij}，在本例中就是

$$R_{ij}=u_{i+1,j}+u_{i,j+1}+u_{i-1,j}+u_{i,j-1}-4u_{ij}+2h^2.\tag{33.3}$$

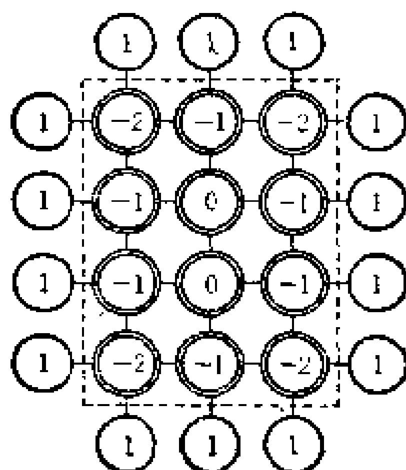
(在 Laplace 方程的情形右边沒有 $2h^2$ 一項。) 將此 R 的值記入网点的右上方(參看图 33.3)。

程序 4　將尽可能使 R 成为 0 的校正值記入左上方，并且追記在同一点右上方的 R 乃至邻近点的 R 由此而产生的变化。图

33.1(a) 是在中心点作了 $+1$ 的校正时，在影响所及的各点把 R 的变化記入 ○ 号內的图示。



(a) 殘差的变化



(b) 分块松弛法之例

图 33.1

图 33.1(b) 是在虛綫內的各点同时作了 $+1$ 的校正时 R 的变化的图示。这样的操作方法叫做分块松弛法 (block relaxation)。为了作分块松弛法，計算工作者考虑了种种的分块形状。

程序 5　依次在各网点进行上述的計算，直到在所有点上的 u 的值都在 $R=\pm2$ 以下的范圍內各自收斂于某个值，此时，关于这样大小的网格的計算认为已經完成，而将网格縮小一半。

程序 6　关于細分后的网格重复上述的計算。

这样进行下去，直到对应于相邻两个 h 值的 u 值，大体上一致到所希望的位数，計算就认为已經完成。或者，在将算子 K 作用于已出現的网点上的 u 的值而求出 R，那么 $-R/20$ 表示使用精确度高于 H 的 K 时的校正值。这是誤差的一个估計。

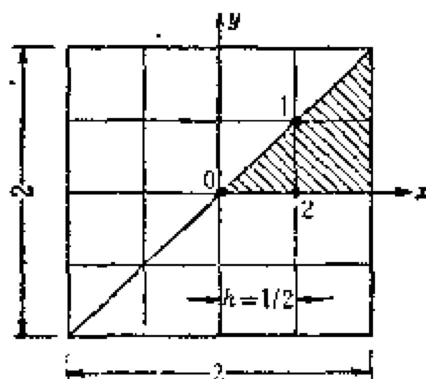


图　33.2

以下用这一程序計算目前的例子。

首先令 $h=\frac{1}{2}$，由图 33.2 看出，依据对称性，只要在图中画有斜綫的部分求解就可以了。(設 $h=1$，那么显然，在中心的 u 的值为 0.50。)

关于 0, 1, 2 这 3 点作 (33.2)，矩陣的形状如下:

$$\begin{pmatrix} -4 & 2 & 0 \\ 2 & -4 & 1 \\ 0 & 4 & -4 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_0 \end{pmatrix} = -\frac{1}{4} \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} \tag{33.4}$$

(其中已按对称性作了变形，因此左边的矩陣已不是对称的了。)

(33.4) 也可以用松弛法求解。但因它是 3 元联立方程組，因此(按程序 2′) 很容易解得

$$u_0 = 0.5625, \quad u_1 = 0.3438, \quad u_2 = 0.4375. \tag{33.5}$$

将这些值在小数第 3 位作舍入而后 1000 倍起来得

$$u_0 = 560, \quad u_1 = 340, \quad u_2 = 440. \tag{33.6}$$

用这些値作为关于上面网格一半的 $h = \frac{1}{4}$ 时的最初几个値，記入图 33.3(a) 中对应的位置。我們可以这些值为基础，按比例关系推出在其他点的值，但也可以由物理的考慮出发，画出連接上面的値和边界值的簡单曲綫而定出預定値。这些值就是图中附有"——"号的値。( ) 中的值是由以上的两种值利用 (33.2) 求得的。最后就得到图 33.3(a) 中网点下的預定値。

首先，以預定値为基础求出 R，就得到在图 (a) 中紧靠着各网点的右上方处記入的値 (R_a)。由此看出，正的 R 較多。此时就应用分块松弛法。在目前的情形，对图中画有虛綫和斜綫部分的整个内部进行分块松弛，其校正値 C_a 是 15，記入紧靠着对应点的左上方。由此产生的殘差 R_b 記入各点右上方 R_a 的上面。以下就关于各点实行松弛法。各个校正值按由下向上的順序記入网点的左上方，而在右上方記入最后的 R。R 在 ± 1 以内。最后的結果是由預定值和校正値的代数和求得的，記入图 (b) 网点的下方。图 (b) 中网点上附有 ( ) 号的値是由本例的解析解

$$u = \frac{32}{\pi^3} \cdot 1000 \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^3} \left\{ 1 - \frac{\operatorname{ch}(2n+1)\pi y/2}{\operatorname{ch}(2n+1)\pi/2} \right\} \cos(2n+1)\frac{\pi x}{2} \tag{33.7}$$

算得的值。誤差約在 2% 以下。

在这里将 K 作用于誤差最大的点 a，得到 $R_a = 43$. 因此，校正值是 $43/20 \approx 2$. 也就是說，使用 K 时，在点 a 的値 142 应增大 2 左右。这个値也可以作为大略的誤差估計。但精确的誤差应該再将网格細分一次进行計算。

如果要进一步提高精确度，需要将两格再縮小一半而使用松弛法。但在这里将叙述另外的方法。

L. Fox 着眼于用差分方程代换时的截断誤差，通过对这一誤差的校正，

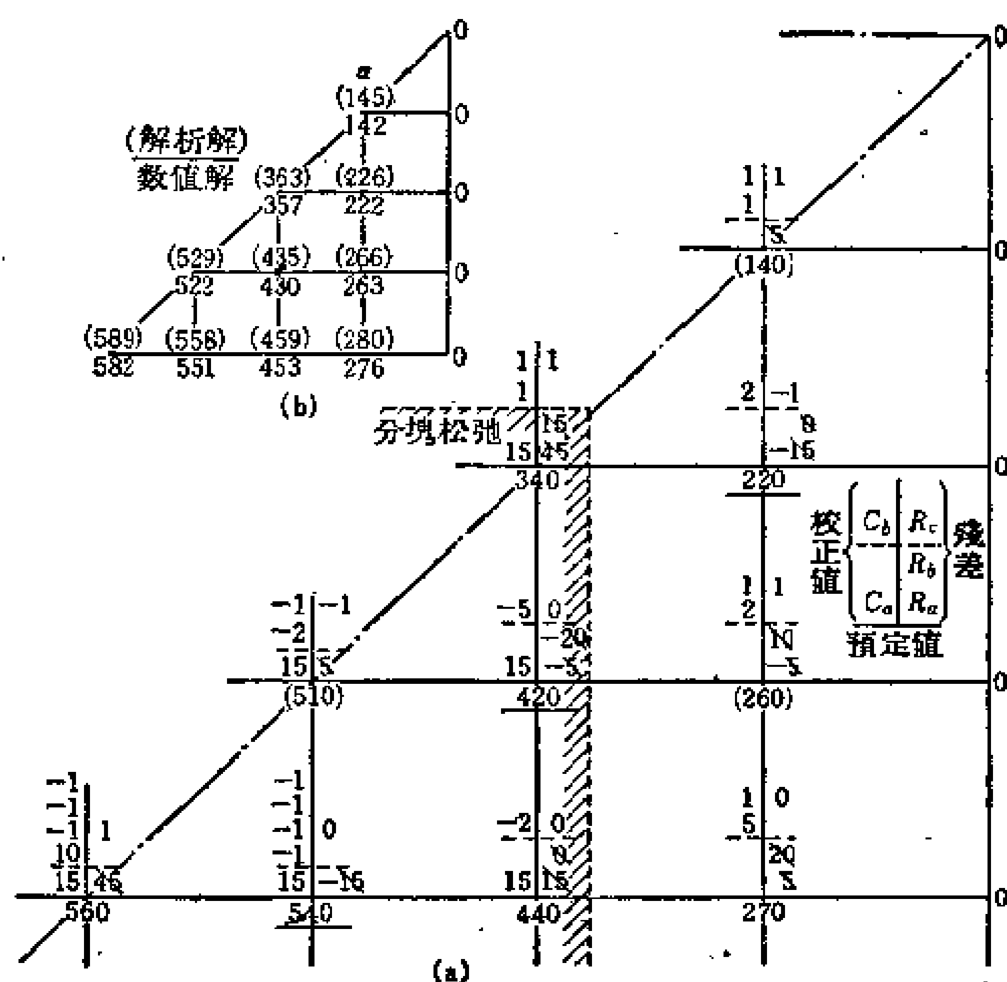


图 33.3

C_a：由分块松弛所得的 } 校正值
C_b：在每个点的
R_a：预定值的 } 残差
R_b：由于作分块松弛的
R_c：最后的

提出了用較少的网格得到比較好的精确度的方法❶，其原理与§23末所述的方法是相同的。以下就目前的例子，对于 $h=1/2$ 的情形进行計算。

考慮截断誤差 Δ，$\boldsymbol{Hu}=-2h^2$ 的式子是

$$\left.\begin{aligned}&u_{i+1,j}+u_{i,j+1}+u_{i-1,j}+u_{i,j-1}-4u_{ij}+2h^2+\Delta u_{ij}=0,\\&\Delta u_{ij}=-\frac{1}{12}(\delta_x^{\mathrm{IV}}+\delta_y^{\mathrm{IV}})u_{ij}+\frac{1}{90}(\delta_x^{\mathrm{VI}}+\delta_y^{\mathrm{VI}})u_{ij}+\cdots.\end{aligned}\right\}\tag{33.8}$$

❶ L. Fox: Proc. Roy. Soc., A 190 (1947), pp. 31~59.

将这分成第 1 近似 $u^{(0)}$，第 1 次校正值 $u^{(1)}$，第 2 次校正值 $u^{(2)}$，… 时，

$$
\left.\begin{aligned}
&u=u^{(0)}+u^{(1)}+u^{(2)}+\cdots,\\
&u^{(0)}_{i+1,j}+u^{(0)}_{i,j+1}+u^{(0)}_{i-1,j}+u^{(0)}_{i,j-1}-4u^{(0)}_{ij}+2h^2=0,\\
&u^{(1)}_{i+1,j}+u^{(1)}_{i,j+1}+u^{(1)}_{i-1,j}+u^{(1)}_{i,j-1}-4u^{(1)}_{ij}+\Delta u^{(0)}_{ij}=0,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots.
\end{aligned}\right\}\qquad(33.9)
$$

一般說来，設 u 在边界上的值为 b（在目前 $b=0$），則 $u^{(0)}$ 和 $u^{(k)}$ ($k=1, 2, \cdots$) 在边界上的値可以考慮作

$$u^{(0)}=b,\quad u^{(1)}=u^{(2)}=\cdots=0.$$

由得到的第 1 近似 $u^{(0)}$ 作差分表，計算 $\delta^{\rm IV}$，加上校正值。如果只在正方形的中心取一个网点，那么 $u_0=500$，此时不能作出差分表，設 $h=1/2$，那么，除去边界附近的点外，可以作出差分表（使用将 (33.5) 的值四舍五入而取至第 3 位的値）。但是，在这边界附近的点，必須使用外插法。方法之一是用某些方法外插 δ^2 的值。这里，在边界上的微分方程中，已經知道一个 2 阶偏导数的值（在目前的情形是 0），因此，另一个 2 阶偏导数的值也是已知的，由此，δ^2 也是已知的。在表 33.1 中用（ ）表示。例如，关于平行于 y 軸的边界，将 u 乘以 10^3，

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=\frac{\partial^2 u}{\partial x^2}=-2000,$$

$$\delta_x^2 u\approx h^2\frac{\partial^2 u}{\partial x^2}=\frac{1}{4}\frac{\partial^2 u}{\partial x^2}.\quad \therefore\quad \delta_x^2 u=-500.$$

由所得的差分表求各点的 Δ 得到

$$\Delta u_0=(-126)\times 2\div 12=-21,$$

$$\Delta u_1=(-188)\times 2\div 12=-32,$$

表 33.1

u	δ_x	δ_x^2	δ_x^3	δ_x^4	u	δ_y	δ_y^2	δ_y^3	δ_y^4
0		(−500)			0		(−500)		
	438		(187)			344		(250)	
438		−313		(−124)	344		−250		(−188)
	125		63			94		62	
563		−250		−126	438		−188		−124
	−125		−63			−94		−62	
438		−313		(−124)	344		−250		(−188)
	−438		(−187)			−344		(−250)	
0		(−500)			0		(−500)		

$$\Delta u_2 = (-124-126) \div 12 = -21.$$

$$\therefore\quad \begin{pmatrix} -4 & 2 & 0 \\ 2 & -4 & 1 \\ 0 & 4 & -4 \end{pmatrix} \begin{pmatrix} u_1^{(1)} \\ u_2^{(1)} \\ u_3^{(1)} \end{pmatrix} = \begin{pmatrix} -32 \\ -21 \\ -21 \end{pmatrix}.$$

由此解得(也可以用松弛法)

$$u_0^{(1)} \approx 26,\quad u_1^{(1)} \approx 19,\quad u_2^{(1)} \approx 21.$$

$$\therefore\quad u_0 = 563+26 = 589,\quad u_1 = 344+19 = 363,\quad u_2 = 438+21 = 459.$$

这一結果較之令 $h=\frac{1}{4}$ 而用松弛法求得的結果精确度要好得多，与解析解一致。

［注1］ 用松弛法进行計算时，有时需要将校正值全部相加重新計算 R 进行驗算。

［注2］ 可以考慮的誤差有两种。一是根源于用差分方程代换微分方程，卽用 H 或 K 代换 ∇^2 而产生的誤差，一是解差分方程时所产生的誤差。关于这些問題請参看 Milne [14] p. 216。又，关于用差分近似解边界值問題时的稳定性，收斂性，誤差估計的理論在 [19] § 37 以下有詳細的討論。

第7章　特征值問題的数値解法

§34 特征值問題

以工程上的例子来說，振动系統的固有振动数，彈性稳定問題中的翹曲載荷，或者原子反应堆为了使連鎖反应持續进行所必需的临界大小等，这些决定某一系統的某些临界条件的参数的值叫做特征值 (eigenvalue)．决定这些特征值的所謂特征值問題，在离散系統的情形，多半归結为求矩陣的特征值的形式，在連續系統的情形則表現为常微分方程、偏微分方程或积分方程的特征值問題的形式。但是，在用数值方法求解时，后者也几乎全部可以近似地归結为求矩陣的特征值的問題。因此，以下主要考虑矩陣的特征值問題，而用如下的极为简单的例子予以說明。

例　求如图 34.1 所示的扭曲振动系的振动数及振动方式。

設旋轉圓板的轉动慣量为 I, $2I$, $4I$，軸的扭曲强度为 $4k$, $2k$, k，各个圆板的旋轉角位移为 θ_1, θ_2, θ_3，用 t 表示时刻，那么这一系統的运动方程为

$$\left.\begin{aligned}
-I\frac{d^2\theta_1}{dt^2}&=4k\theta_1-2k(\theta_2-\theta_1),\\
-2I\frac{d^2\theta_2}{dt^2}&=2k(\theta_2-\theta_1)-k(\theta_3-\theta_2),\\
-3I\frac{d^2\theta_3}{dt^2}&=k(\theta_3-\theta_2).
\end{aligned}\right\}\tag{34.1}$$

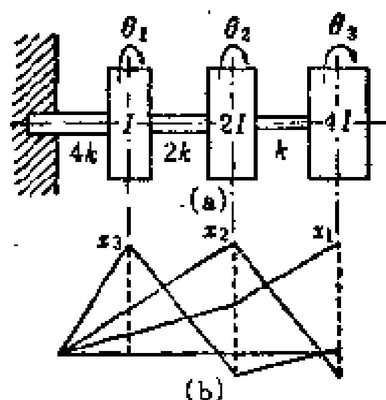


图　34.1

假定振动的形状是 $\theta_i=x_i\sin(\omega t+\varphi)$ (x_i 表示振幅，ω 表示角頻率，φ 表示初相)，那么 (34.1) 就成为

$$\boldsymbol{Ax}=\lambda\boldsymbol{Bx},\tag{34.2}$$

$$A=\begin{pmatrix} 6 & -2 & 0 \\ -2 & 3 & -1 \\ 0 & -1 & 1 \end{pmatrix},\quad B=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix},$$

$$x=\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \tag{34.3}$$

其中 $\lambda=I\omega^2/k$。

在通常出現的工程或物理学問題中，如上例中那样，A, B 都是对称的，而且至少其中之一是正定形式。这里举的也是这种类型的例子。現在設 B 为正定形式，那么 B^{-1} 存在，因而可以把(34.2)改写成

$$Hx=\lambda x, \tag{34.4}$$

其中

$$H=B^{-1}A.$$

在目前的情形，可以簡单地求得 H 为

$$H=\begin{pmatrix} 6 & -2 & 0 \\ -1 & 3/2 & -1/2 \\ 0 & -1/4 & 1/4 \end{pmatrix}, \tag{34.5}$$

但是，在一般情形，設矩陣的阶数为 n，那么为求 $B^{-1}A$，需要作大約相当于 n^3 次乘法的計算，这不是很容易的。而且，这里 H 已經不是对称的了。从(34.2)或(34.4)的形状考虑，这是矩陣的特征值問題❶。这样的齐次方程組具有非零解 x 的必要条件是

$$|\lambda B-A|=0, \tag{34.6}$$

或

$$D(\lambda)\equiv|\lambda I-H|=0. \tag{34.7}$$

展开方程(34.6)或(34.7)的左边，就得到关于 λ 的 n 次代数方程。

❶ 在§3的例中我們已經遇到过这种特征值問題之一。

通常把这方程叫做特征方程，在振动問題中特别叫做振动数方程。λ 可以作为这两个方程中任意一个的根而得到，只有在 λ 取这些值时，向量 $\boldsymbol{x}$ 才能取非 0 值。这 λ 就是特征值，而向量 $\boldsymbol{x}$ 就是特征向量(eigenvector)。

在本例中，展开(34.7)的行列式就得到

$$\lambda^3-(31/4)\lambda^2+(35/4)\lambda-1=0. \tag{34.8}$$

它的根，即特征值是

$$\lambda_1=0.128716,\quad \lambda_2=1.212186,\quad \lambda_3=6.409098 \tag{34.9}$$

(由此就可以得到振动数)。对应的特征向量，可以对每个 λ_j 計算矩陣 $(\lambda_j\boldsymbol{I}-\boldsymbol{H})$ 中某一行的余因子之間的比而求得如下：

$$\boldsymbol{x}_1=\begin{pmatrix}0.165257\\0.485136\\1.000000\end{pmatrix},\quad \boldsymbol{x}_2=\begin{pmatrix}0.417727\\1.000000\\-0.259825\end{pmatrix},\quad \boldsymbol{x}_3=\begin{pmatrix}1.000000\\-0.204549\\0.008302\end{pmatrix}. \tag{34.10}$$

在本例考察的系統中，$\boldsymbol{x}_j$ 的分量之比表示对于 λ_j (因而对于第 j 次振动数)的各圓板的振幅之比(图 34.1(b))。

象在本例中这样，n 的值很小的情形，使用初等的行列式展开法也不需要費很大的劳力，但若 n 等于 4,5 或 6 就不那么容易了。本章的目的就在于說明几个当 n 的值很大时也能用机械进行計算的方法。

求矩陣的特征值、特征向量的方法，大体上可分为两种。一种是如上面所作的那样，化成(34.6)或(34.7)的形状而后展开，导出特征方程的直接方法。其中的展开法又有好几种。另一种是，或者把試解向量迭次代入而进行改善，或者改善矩陣使它对角綫化的間接方法。

§35 直 接 方 法

直接方法就是把 $D(\lambda)=|\lambda\boldsymbol{I}-\boldsymbol{H}|$ 直接展开，作为 λ 的 n 次代数方程的根而求特征值的方法，因而一下就可以求得全部的特征

值。展开的方法很多，这里介紹用易于利用到的 IBM 602A 計算穿孔机也能容易地进行的 Frame 方法。

原理請参看 Dwyer[1] p. 225，这里只叙述操作方法。

考虑(34.7)的展开形式

$$D(\lambda)\equiv\lambda^n-c_1\lambda^{n-1}-c_2\lambda^{n-2}-\cdots-c_{n-1}\lambda-c_n=0. \tag{35.1}$$

由于 $\lambda\boldsymbol{I}-\boldsymbol{H}$ 的伴随矩陣关于 λ 的次数不高于 $(n-1)$，因此

$$\begin{aligned}C(\lambda)&\equiv\mathrm{adj}(\lambda\boldsymbol{I}-\boldsymbol{H})\\&=\lambda^{n-1}\boldsymbol{H}_0+\lambda^{n-2}\boldsymbol{H}_1+\cdots+\lambda^{n-k}\boldsymbol{H}_{k-1}+\cdots+\boldsymbol{H}_{n-1}.\end{aligned} \tag{35.2}$$

以上两式中的 c_k 是未定系数而 $\boldsymbol{H}_k$ 是未定方陣，这些未定量可以由下列的公式求得：

$$\left.\begin{array}{ll}\boldsymbol{H}_0=\boldsymbol{I}, & c_1=\mathrm{trace}(\boldsymbol{H}\boldsymbol{H}_0)=\mathrm{trace}(\boldsymbol{H}),\\ \boldsymbol{H}_1=\boldsymbol{H}\boldsymbol{H}_0-c_1\boldsymbol{I}, & c_2=\mathrm{trace}(\boldsymbol{H}\boldsymbol{H}_1)/2,\\ \boldsymbol{H}_2=\boldsymbol{H}\boldsymbol{H}_1-c_2\boldsymbol{I}, & c_3=\mathrm{trace}(\boldsymbol{H}\boldsymbol{H}_2)/3,\\ \cdots\cdots, & \cdots\cdots,\\ \boldsymbol{H}_k=\boldsymbol{H}\boldsymbol{H}_{k-1}-c_k\boldsymbol{I}, & c_{k+1}=\mathrm{trace}(\boldsymbol{H}\boldsymbol{H}_k)/(k+1),\\ \cdots\cdots, & \cdots\cdots,\end{array}\right\} \tag{35.3}$$

其中 $\mathrm{trace}(\boldsymbol{H}\boldsymbol{H}_k)$ 表示矩陣 $\boldsymbol{H}\boldsymbol{H}_k$ 的迹，即主对角綫元素之和。

例　以下按照公式(35.3)展开前面的例子。

$$c_1=6+\frac{3}{2}+\frac{1}{4}=\frac{31}{4},$$

$$\boldsymbol{H}_1=\boldsymbol{H}\boldsymbol{H}_0-c_1\boldsymbol{I}=\begin{pmatrix}6-\frac{31}{4} & -2 & 0\\ -1 & \frac{3}{2}-\frac{31}{4} & -\frac{1}{2}\\ 0 & -\frac{1}{4} & \frac{1}{4}-\frac{31}{4}\end{pmatrix}$$

$$=\begin{pmatrix}-\frac{7}{4} & -2 & 0\\ -1 & -\frac{25}{4} & -\frac{1}{2}\\ 0 & -\frac{1}{4} & -\frac{15}{2}\end{pmatrix},$$

$$HH_1=\begin{pmatrix} -\dfrac{17}{2} & \dfrac{1}{2} & 1 \\ \dfrac{1}{4} & -\dfrac{29}{4} & 3 \\ \dfrac{1}{4} & \dfrac{3}{2} & -\dfrac{7}{4} \end{pmatrix},$$

$$\text{trace}\, HH_1=-\frac{17}{2}-\frac{29}{4}-\frac{7}{4}=-\frac{35}{2}.$$

$$\therefore \quad c_2=-\frac{35}{4}.$$

类似地求出 H_2 而作 HH_2，就得到 $HH_2=I$. ∴ $\text{trace}\, HH_2=3$, $c_3=1$. 因此 $H_3=0$. 这起了驗算的作用。

由以上的計算得到

$$\lambda^3-\frac{31}{4}\lambda^2+\frac{35}{4}\lambda-1=0,$$

与前面得到的(34.8)相同，又

$$C(\lambda)=\lambda^2H_0+\lambda H_1+H_2$$

$$=\begin{pmatrix} \lambda^2-\dfrac{7}{4}\lambda+\dfrac{1}{4} & -2\lambda+\dfrac{1}{2} & 1 \\ -\lambda+\dfrac{1}{4} & \lambda^2-\dfrac{25}{4}\lambda+\dfrac{3}{2} & -\dfrac{1}{2}\lambda+3 \\ \dfrac{1}{4} & -\dfrac{1}{4}\lambda+\dfrac{3}{2} & \lambda^2-\dfrac{15}{2}\lambda+7 \end{pmatrix}. \qquad (35.4)$$

在此代入 λ_j，由伴隨矩陣的性质推出，各列表示特征向量 x_j 的分量之比。例如，对于 $\lambda_3=6.409098$，

$$C(6.409098)=\begin{pmatrix} 30.110616 & -12.318196 & 1.000000 \\ -6.159098 & 2.519675 & -0.204549 \\ 0.250000 & -0.102274 & 0.008302 \end{pmatrix}.$$

不論那一列的比都是相同的，等于最右一列之比，而且与(34.10)所示的結果相同。由此可見，Frame 方法有由 $C(\lambda)$ 决定特征向量的便利之点。

直接方法中，除了 Frame 方法外，还有間接展开法、trace 法或 Hessenberg 方法等。其中除 Hessenberg 方法只需要比例于 n^3 的計算量外，其他方法都需要比例于 n^4 的計算量，当 n 的值較大时即成为龐大的計算量（这些計算量当然都不包含解代数方程

的計算在內)。

§36 迭代法(1)

这里叙述把試解向量逐步改善的迭代法,这是間接方法之一。根据这方法,在求得一个特征向量的同时,也求得了对应于它的特征值。但其他的特征值、特征向量必須另外利用特征向量的正交性等性质才能求得。虽然如此,操作却比較簡单,用 IBM 602A 計算穿孔机一类的計算机就可以进行計算。这方法的理論是比較古典的,1885年已經由 H. A. Schwarz 給出,常用于工程上振动、翹曲等問題中。

就(34.4)的形状加以考虑。关系式(34.4)表示,将 $\boldsymbol{H}$ 作用于特征向量 $\boldsymbol{x}$ 的結果就等于向量 $\boldsymbol{x}$ 的各分量乘以某一純量(即特征值 λ)的結果。如果代替 $\boldsymbol{x}$,将 $\boldsymbol{H}$ 作用于試解向量 $\boldsymbol{v}_0$,所得結果 $\boldsymbol{v}_1$ 与 $\boldsymbol{v}_0$ 是不同的,$\boldsymbol{v}_0$ 与 $\boldsymbol{v}_1$ 的各分量之間也沒有一定的比例关系。为了考察这一比例关系,在一开始时就設 $\boldsymbol{v}_0$ 的最大分量为1,然后提出純量 λ,使 $\boldsymbol{v}_1$ 的对应分量也等于1. 其次,再把 $\boldsymbol{H}$ 作用于 $\boldsymbol{v}_1$,求出 $\boldsymbol{v}_2$,实行同上的手續。这样逐次反复进行。

例1　对(34.4)作上述的計算。

設 $\boldsymbol{v}_0=(1, 0, 0)'$,作上述的計算,得到如下的一系列結果:

$$\left.\begin{aligned}
\boldsymbol{H}\boldsymbol{v}_0&=6(1, -1/6, 0)',\\
\boldsymbol{H}\boldsymbol{v}_1&=6.333(1, -0.1974, 0.0066)',\\
\boldsymbol{H}\boldsymbol{v}_2&=6.3947(1, -0.2032, 0.0080)',\\
\boldsymbol{H}\boldsymbol{v}_3&=6.4064(1, -0.2043, 0.0082)',\\
\boldsymbol{H}\boldsymbol{v}_4&=6.4086(1, -0.2045, 0.0083)',\\
\boldsymbol{H}\boldsymbol{v}_5&=6.4090(1, -0.2045, 0.0083)',\\
\boldsymbol{H}\boldsymbol{v}_6&=6.4091(1, -0.20455, 0.00830)'.
\end{aligned}\right\}\tag{36.1}$$

这样,相邻接的 $\boldsymbol{v}$ 的分量之比几乎收斂于一定值 6.4091. 这个値等于前面已經求得的最大特征值 λ_3,而对应的向量 $\boldsymbol{v}$ 則等于(34.10)的 $\boldsymbol{x}_3$. 上述的理由可以簡单地叙述如下。

一般說来,任意向量可以按特征向量展开。設 $\boldsymbol{v}_0$ 可以展开成

$$\boldsymbol{v}_0=\sum_{j=1}^{n} c_j\boldsymbol{x}_j. \tag{36.2}$$

那么,

$$\left.\begin{aligned}\boldsymbol{v}_1&=\boldsymbol{H}\boldsymbol{v}_0=\sum_{j=1}^{n} c_j\lambda_j\boldsymbol{x}_j,\\ \boldsymbol{v}_2&=\boldsymbol{H}\boldsymbol{v}_1=\sum_{j=1}^{n} c_j\lambda_j^2\boldsymbol{x}_j,\\ &\cdots\cdots,\\ \boldsymbol{v}_k&=\boldsymbol{H}\boldsymbol{v}_{k-1}=\sum_{j=1}^{n} c_j\lambda_j^k\boldsymbol{x}_j.\end{aligned}\right\} \tag{36.3}$$

这里省去了提出純量的手續,但在实质上并无差别。

但是,如果設最大特征値为 λ_n,而且 $0<\lambda_1<\lambda_2<\cdots<\lambda_n$,并設 $c_n\neq 0$,那么

$$\boldsymbol{v}_k=c_n(\lambda_n)^k\left[\boldsymbol{x}_n+\sum_{j=1}^{n-1}\frac{c_j}{c_n}\left(\frac{\lambda_j}{\lambda_n}\right)^k\boldsymbol{x}_j\right]. \tag{36.4}$$

因为 $|\lambda_j/\lambda_n|<1\,(j\neq n)$,因此,当反复計算的次数 k 增大时,上式中求和的項($\sum$ 項)和 $\boldsymbol{x}_n$ 項比較起来可以忽略不計。由此可見 $\boldsymbol{v}_k$ 与 $\boldsymbol{v}_{k-1}$ 的分量之比向 λ_n 收斂,而且 $\boldsymbol{v}_k$ 收斂于 $\boldsymbol{x}_n$ 的純量倍。

这样就可以看到,在迭代法中,逐次得到的向量序列,向对应于最大特征値 λ_n 的特征向量 $\boldsymbol{x}_n$ 收斂。显然,当 $|c_j/c_n|\,(j\neq n)$, $|\lambda_j/\lambda_n|\,(j\neq n)$ 两者愈小时,到达收斂所需要的反复次数 k 也愈小。前者表示,应該选取 $\boldsymbol{v}_0$ 尽可能地接近 $\boldsymbol{x}_n$ 的形状,而后者則由矩陣的性质决定。如果在 $\boldsymbol{v}_0$ 的表示式(36.2)中設 $c_n=0$,那么,理論上应該在 λ_n 的其次得到最大特征値,但在实际上,由于舍入的影响,$\boldsymbol{x}_n$ 的分量仍然微小地出現,而后逐步扩大,結果仍向 $\boldsymbol{x}_n$ 收斂。在振动或翘曲問題中,考虑到可能发生振动或形变的形式,而选取与此接近的 $\boldsymbol{v}_0$ 就可以了。本例中的 $\boldsymbol{v}_0$ 也就是由这种考虑选取的。

关于 $|\lambda_j/\lambda_n|$,由于它决定于矩陣的性质,很难予以左右,但如

这比值很小，接近于1/10左右，那么，每一次代入可以提高精确度1位左右，但如这比值接近于1，那么收斂就非常之慢。求出H^2，H^3等的特征值而用差分法求解也是一个方法，但与此相比較，用作出矩陣的适当的多項式（下节），或其他方法要更好一些。

在振动問題中，要求λ的最小值（最低次固有振动数）。此时，只要求出$1/\lambda$的最大值就可以了，因此，設

$$H^{-1}x=(1/\lambda)x, \tag{36.5}$$

而考虑为求$1/\lambda$的特征值的問題。H^{-1}可以作为H的逆矩陣而求得，但在工程問題中，有时也可由系統的性质直接求得。

例2　在目前的例中，可以由系統的彈性性质作为影响系数而求得，对于这样得到的

$$H^{-1}=\begin{pmatrix}\frac{1}{4} & \frac{1}{2} & 1\\ \frac{1}{4} & \frac{3}{2} & 3\\ \frac{1}{4} & \frac{3}{2} & 7\end{pmatrix}, \tag{36.6}$$

設$v_0=(0,0,1)'$而开始計算，在第3～6次的迭代中得到如下的結果：

$$\left.\begin{aligned}H^{-1}v_2&=7.759(0.1650,\ 0.4845,\ 1.0000),\\ H^{-1}v_3&=7.768(0.1652,\ 0.4851,\ 1.0000),\\ H^{-1}v_4&=7.768906(0.165254,\ 0.485127,\ 1.000000)',\\ H^{-1}v_5&=7.769004(0.165256,\ 0.485134,\ 1.000000)'.\end{aligned}\right\} \tag{36.7}$$

式中的向量与x_1几乎完全一致，而$1/7.76900\approx0.128717$等于λ_1。

如上所述，就能够求出最大或最小特征值。以下討論中間特征值的求法。

§37　中間特征值的求法

求最大特征值与最小特征值以外的中間特征值，有利用正交条件的方法和利用矩陣多項式的方法。

1）利用正交条件的方法

例1　在§34的例中，計算$x_3'Bx_2$。

利用(34.3)，(34.10)得到

$$x_3' B x_2 = (1.0000,\ -0.204549,\ 0.008302)\begin{pmatrix}1 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 4\end{pmatrix}\begin{pmatrix}0.165257\\ 0.485136\\ 1.000000\end{pmatrix}$$

$$\approx 0. \tag{37.1}$$

这就是說，特征向量 $\boldsymbol{x}_3$ 与 $\boldsymbol{x}_2$ 关于矩陣 $\boldsymbol{B}$ 是正交的。一般說来，在关于两个相异特征値的特征向量之間，关系式

$$\boldsymbol{x}_i' \boldsymbol{B} \boldsymbol{x}_j = 0 \quad (i \neq j) \tag{37.2}$$

成立。已知最大特征値时，由在正交条件(37.2)中令 $i=n$ 所得的式子和 $\boldsymbol{H}\boldsymbol{x}_j=\lambda\boldsymbol{x}_j$，可以求得 $\boldsymbol{x}_n$ 以外的諸向量。

例 2 設 x_2 的分量为 $x_1^{(2)}$, $x_2^{(2)}$, $x_3^{(2)}$，在(37.1)中用这些分量代 x_2 得到

$$x_1^{(2)} - 0.409098 x_2^{(2)} + 0.033288 x_3^{(2)} = 0. \tag{37.3}$$

由此将 $x_1^{(2)}$ 用 $x_2^{(2)}$, $x_3^{(2)}$ 表示出而代入 $\boldsymbol{H}\boldsymbol{x}_j=\lambda\boldsymbol{x}_j$ 的 $x_1^{(2)}$ 得到

$$\left.\begin{aligned} 0.454588 x_2^{(2)} - 0.199728 x_3^{(2)} &= \lambda x_1^{(2)},\\ 1.090902 x_2^{(2)} - 0.466712 x_3^{(2)} &= \lambda x_2^{(2)},\\ -0.25\ x_2^{(2)} + 0.25\ x_3^{(2)} &= \lambda x_3^{(2)}. \end{aligned}\right\} \tag{37.4}$$

利用(37.4)的后两式，用矩陣

$$\boldsymbol{H}_2 = \begin{pmatrix} 1.090902 & -0.466712\\ -0.25 & 0.25 \end{pmatrix}$$

进行迭代就可以了。設 $\boldsymbol{v}_0=(1,\ 0)'$ 而进行計算，逐步得到

$$\begin{pmatrix}1\\0\end{pmatrix},\ \overset{1.90902}{\begin{pmatrix}1.0\\-0.22917\end{pmatrix}},\ \overset{1.19786}{\begin{pmatrix}1.0\\-0.25634\end{pmatrix}},\ \overset{1.210539}{\begin{pmatrix}1.0\\-0.259459\end{pmatrix}},$$

$$\overset{1.211995}{\begin{pmatrix}1.0\\-0.259791\end{pmatrix}},\ \overset{1.212150}{\begin{pmatrix}1.0\\-0.259826\end{pmatrix}}.$$

每个向量上方的数是提出来的純量，随着迭代的进行这純量逐步接近第2特征值 λ_2。而对应的向量逐步接近 x_2 的第2，第3分量 $x_2^{(2)}$ 和 $x_3^{(2)}$。至于第1分量 $x_1^{(2)}$，可以由(37.3)求得如下：

$$x_1^{(2)} = 0.409098(1.0) - 0.033288(-0.259826) \approx 0.417747.$$

其次，关于 x_1 有如下的两个正交条件：

$$\left.\begin{aligned} x_1^{(1)}-0.409098x_2^{(1)}+0.033288x_3^{(1)}=0, \\ 0.417747x_1^{(1)}+2x_2^{(1)}-1.039304x_3^{(1)}=0. \end{aligned}\right\} \quad (37.5)$$

由此将 $x_1^{(1)}$, $x_2^{(1)}$ 用 $x_3^{(1)}$ 表示出，得到

$$x_1^{(1)}=0.165168x_3^{(1)}, \quad x_2^{(1)}=0.485149x_3^{(1)}. \quad (37.6)$$

这些值与由(34.10)給出的 x_1 几乎相等。

这方法的缺点在于，利用正交性逐步求特征向量的过程中，由于舍入的原因而逐步丧失精确度。因此，在詳細地求解时，必須把特征向量的分量計算到所需要的精确度以上。从而，用这迭代法求与最大特征值邻接的1～2个特征值，或求与最小特征值邻接的1～2个特征值，以及对应的特征向量时是可以使用的，但如用于进一步的求解，其計算就越来越困难了。

2) **用矩陣多項式的方法**

这方法是用矩陣 $\boldsymbol{H}$ 的多項式作为算子而代替 $\boldsymbol{H}$ 的方法，其原理如下。

在(34.4)的两边由左方作用 $\boldsymbol{H}$ 得到

$$\boldsymbol{H}^2\boldsymbol{x}=\lambda\boldsymbol{H}\boldsymbol{x}=\lambda^2\boldsymbol{x}.$$

这样反复进行下去得到

$$\boldsymbol{H}^k\boldsymbol{x}=\lambda^k\boldsymbol{x} \quad (k=1, 2, \cdots). \quad (37.7)$$

这样得到的 $\boldsymbol{H}^k$ 的特征向量与 $\boldsymbol{H}$ 的特征向量相同，而其特征值是 $\boldsymbol{H}$ 的对应特征值的 k 次幂。对于多項式 $a_k\boldsymbol{H}^k+\cdots+a_1\boldsymbol{H}+a_0\boldsymbol{I}$ 也可以得到类似的結論。即，关系式

$$(a_k\boldsymbol{H}^k+\cdots+a_1\boldsymbol{H}+a_0\boldsymbol{I})\boldsymbol{x}=(a_k\lambda^k+\cdots+a_1\lambda+a_0)\boldsymbol{x} \quad (37.8)$$

成立。(左边算子的)特征向量与 $\boldsymbol{H}$ 的特征向量是相同的，但特征值是由 $\boldsymbol{H}$ 的对应特征值的多項式构成的。由此，作出在所求的特征值处取得极大值的多項式，就可以使这特征值轉化为最大特征

值❶。用于这一目的的多項式中,最簡单的形状是 $\boldsymbol{H}+a_0\boldsymbol{I}$。

利用这一綫型矩陣应当能够求得目前我們討論的最小特征值。取 $\boldsymbol{H}-4\boldsymbol{I}$ 进行計算,經过 8 次迭代得到 $\lambda=-3.853$,与精确值 $\lambda=0.128716-4=-3.871284$ 相差相当大。由 $(\lambda_2-4)/(\lambda_1-4)\approx0.72$ 也可以預計到其收斂性不佳。

其次假定,用后面叙述的松弛法(例如 § 39 例 2)或其他某种方法已經求得最大特征值与最小特征值的近似值为 $\lambda_1\approx0.13$, $\lambda_3\approx6.4$。那么,可以想見,对应于

$$P(\lambda)=(\lambda-0.13)(\lambda-6.4)=\lambda^2-6.54\lambda+0.832$$

的 $P(\boldsymbol{H})$,对于 $\boldsymbol{H}$ 的特征值 λ_1 和 λ_3,具有較小的特征值,而对于 λ_2,有較大的特征值。将 $P(\lambda)$ 的形状略加改善为

$$P(\lambda)=\lambda^2-\frac{13}{2}\lambda+\frac{5}{6}. \tag{37.9}$$

这样,$P(\boldsymbol{H})$ 就成为

$$P(\boldsymbol{H})=\boldsymbol{H}^2-\frac{13}{2}\boldsymbol{H}+\frac{5}{6}\boldsymbol{I}=\frac{1}{12}\begin{pmatrix}-2 & -24 & 12\\ -12 & -54.5 & 28.5\\ 3 & 14.25 & -7.25\end{pmatrix}. \tag{37.10}$$

在这里,設 $v_0=(0,1,0)'$ 而进行計算,經过 4 次迭代得到

$$-66.91835(0.417725,\ 1.000000,\ -0.259825)',$$

收斂得比較快。令所得的特征值 -66.91835 等于(37.9)就可以求出原来的 λ。但也可以将 H 作用于已求得的特征向量而取各分量的比,或者用后述的 Rayleigh 方法也可以求得。

这一方法的特点在于,只要已知其他某些特征值,例如最大最小特征值的那怕是粗略的近似值,就可以求出某一特定特征值的精确值。至于这里的其他某些特征值,可以用前面叙述过的方法予以估計,也可以用 § 39 所述的松弛法等进行估計。

§ 38 迭代法(2)

§ 36 的方法是就特征向量之一把試解向量逐步改善的方法。在本节中,

---

❶ W. M. Kincaid: Quart. Appl. Math., 5, pp. 320~346 (1947); L. F. Richardson: Trans. Roy. Soc. (London), **A 242**, pp. 439~491 (1950).

簡单說明 C. Lanczos 所提出的方法❶并应用这方法来解前面的例子。这方法与前述的迭代法不同，經过有限次，即 n 次后(如果沒有舍入誤差)就得到精确的特征方程。这方法的基本想法就是共軛斜量法 (§ 4) 中想法的基础，即把試解向量 $\boldsymbol{v}_i$ 逐步改善使它們相互正交。

考慮 $\boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{B}\boldsymbol{x}$, $\boldsymbol{H}=\boldsymbol{B}^{-1}\boldsymbol{A}$ 的情形。取任意向量 $\boldsymbol{v}_0$ 作为第一个試解向量，而計算出

$$\boldsymbol{v}_1^*=\boldsymbol{H}\boldsymbol{v}_0. \tag{38.1}$$

在 § 36 的迭代法中，以下立即就把 $\boldsymbol{v}_1^*$ 用作 $\boldsymbol{v}_1$ 而求 $\boldsymbol{H}\boldsymbol{v}_1$，但在目前的情形，則取和 $\boldsymbol{v}_0$ "B-正交"的向量

$$\boldsymbol{v}_1=\boldsymbol{v}_1^*-\alpha_0\boldsymbol{v}_0, \tag{38.2}$$

$$\alpha_0=(\boldsymbol{v}_0'\boldsymbol{B}\boldsymbol{v}_1^*)/(\boldsymbol{v}_0'\boldsymbol{B}\boldsymbol{v}_0) \tag{38.3}$$

作为 $\boldsymbol{v}_1$. 其次，再用由

$$\boldsymbol{v}_2^*=\boldsymbol{H}\boldsymbol{v}_1 \tag{38.4}$$

得到的 $\boldsymbol{v}_2^*$ 作成与 $\boldsymbol{v}_0$ 和 $\boldsymbol{v}_1$ B-正交的 $\boldsymbol{v}_2$:

$$\boldsymbol{v}_2=\boldsymbol{v}_2^*-\alpha_1\boldsymbol{v}_1-\beta_0\boldsymbol{v}_0, \tag{38.5}$$

$$\alpha_1=(\boldsymbol{v}_1'\boldsymbol{B}\boldsymbol{v}_2^*)/(\boldsymbol{v}_1'\boldsymbol{B}\boldsymbol{v}_1), \quad \beta_0=(\boldsymbol{v}_0'\boldsymbol{B}\boldsymbol{v}_2^*)/\boldsymbol{v}_0'\boldsymbol{B}\boldsymbol{v}_0). \tag{38.6}$$

这样下去，就可以作成相互 B-正交的 n 个向量 $\boldsymbol{v}_0, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_{n-1}$. 一般表示式是

$$\left.\begin{aligned}\boldsymbol{v}_i^*&=\boldsymbol{H}\boldsymbol{v}_{i-1},\\ \boldsymbol{v}_i&=\boldsymbol{v}_i^*-\alpha_{i-1}\boldsymbol{v}_{i-1}-\beta_{i-2}\boldsymbol{v}_{i-2},\\ \alpha_{i-1}&=(\boldsymbol{v}_{i-1}'\boldsymbol{B}\boldsymbol{v}_i^*)/(\boldsymbol{v}_{i-1}'\boldsymbol{B}\boldsymbol{v}_{i-1}),\\ \beta_{i-2}&=(\boldsymbol{v}_{i-2}'\boldsymbol{B}\boldsymbol{v}_i^*)/(\boldsymbol{v}_{i-2}'\boldsymbol{B}\boldsymbol{v}_{i-2}).\end{aligned}\right\} \tag{38.7}$$

因为只能取 n 个相互正交的向量，因此，由上述方法作成的向量 $\boldsymbol{v}_n$ 必然是0。用矩陣多項式写出就是

$$\left.\begin{aligned}&p_0(\boldsymbol{H})=\boldsymbol{I},\\ &p_1(\boldsymbol{H})\boldsymbol{v}_0=(\boldsymbol{H}-\alpha_0\boldsymbol{I})\boldsymbol{v}_0=\boldsymbol{v}_1,\\ &p_2(\boldsymbol{H})\boldsymbol{v}_0=\{(\boldsymbol{H}-\alpha_1\boldsymbol{I})p_1(\boldsymbol{H})-\beta_0\boldsymbol{I}\}\boldsymbol{v}_0=\boldsymbol{v}_2,\\ &\quad\cdots\cdots,\\ &p_n(\boldsymbol{H})\boldsymbol{v}_0=\cdots \qquad\qquad =\boldsymbol{v}_n=0.\end{aligned}\right\} \tag{38.8}$$

如果用 λ 代 $\boldsymbol{H}$，就得到如下的多項式:

---

❶ C. Lanczos: J. Res. Nat. Bur. Standards, 45, pp. 255~282 (1950).

$$\left.\begin{aligned}
&p_0(\lambda)=1,\\
&p_1(\lambda)=\lambda-\alpha_0,\\
&p_2(\lambda)=(\lambda-\alpha_1)p_1(\lambda)-\beta_0p_0(\lambda),\\
&p_3(\lambda)=(\lambda-\alpha_2)p_2(\lambda)-\beta_1p_1(\lambda),\\
&\quad\cdots\cdots,\\
&p_n(\lambda)=(\lambda-\alpha_{n-1})p_{n-1}(\lambda)-\beta_{n-2}p_{n-2}(\lambda)=0.
\end{aligned}\right\}\qquad(38.9)$$

最后一式就是所求矩陣的特征方程。

例 仍就前面的例子进行計算。取 $v_0=(1,0,0)'$.

$$v_1^*=Hv_0=(6,-1,0)',$$
$$\alpha_0=(1,0,0)'B(6,-1,0)/(1,0,0)'B(1,0,0)=6,$$
$$v_1=v_1^*-\alpha_0v_0=(0,-1,0),$$
$$v_2^*=Hv_1=(2,-1.5,0.25).$$

以下按前面的公式逐步求得

$$\alpha_1=3/2=1.5,\quad \beta_0=2/1=2,$$
$$v_2=(0,0,0.25),\quad v_3^*=(0,-0.125,0.0625),$$
$$\alpha_2=0.25,\quad \beta_1=0.125.$$

由此并得 $p_1(\lambda)=\lambda-6,\quad p_2(\lambda)=(\lambda-1.5)(\lambda-6)-2,$

$$p_3(\lambda)=(\lambda-0.25)\{(\lambda-1.5)(\lambda-6)-2\}-0.125(\lambda-6)$$
$$=\lambda^3-(31/4)\lambda^2+(35/4)\lambda-1.$$

而 $p_3(\lambda)$ 就是前面已經求得的特征方程。

上述的計算，如果換算成乘法，計算量共有 $4n^3+(13/2)n^2-n/2$ 次，那是相当复杂的。也可以在計算的中途中止而將所得結果作为近似值。

此外，还有用方陣反复乘矩陣本身而使它变成对角矩陣的方法。这方法具有可以得到全部特征値、特征向量的优点。其理論虽然很古老，但因适用于計算机，近年来开始受到重視。这里只列举它的出处❶。

§39 Rayleigh 商以及其他定理的应用

在本节中，叙述本丛书《变分法及其应用》§69，§70 的应用例子以及与此有关的若干定理的应用❷。

---

❶ R. T. Gregory: Math. Tables and Other Aids to Computation, 7, pp. 215～220 (1953)，又，Crandall [11], pp. 118～122.

❷ 参照本丛书《微分方程的近似解法》§24.

1) **Rayleigh 商**　在(34.2)的两边左乘 $\boldsymbol{x}'$，就 λ 解出得到

$$\lambda=(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})/(\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}). \tag{39.1}$$

如果 $\boldsymbol{x}$ 是特征向量，那么(39.1)应該与对应的特征値相等。如果代替特征向量，用試解向量 $\boldsymbol{v}$ 作成同样形状的式子

$$\lambda_R=(\boldsymbol{v}'\boldsymbol{A}\boldsymbol{v})/(\boldsymbol{v}'\boldsymbol{B}\boldsymbol{v}), \tag{39.2}$$

称之为 Rayleigh 商。当 $\boldsymbol{v}$ 在特征向量 $\boldsymbol{x}$ 附近变化时 λ_R 取逗留值，这逗留值就是特征值。一般說来，λ_R 給出最小特征値的上确界，但如用适当选取的基底向量 $u_1, u_2, \cdots, u_n$ 的1次式 $\xi_1u_1+\cdots+\xi_nu_n$ 作为試解向量 $\boldsymbol{v}$，就得到对应于原来的特征值問題的关于 $(\xi_1, \cdots, \xi_n)$ 的特征值問題，它的特征値分别是原来的对应特征值的上确界(前引《变分法及其应用》§69)。

例1　由(36.7)的 $\boldsymbol{v}_3$，令 $\boldsymbol{v}=(0.165, 0.485, 1.00)$ 而代入(39.2)，得到

$$\lambda_R=0.1287165,$$

与由(34.9)給出的解只有舍入誤差程度的差。如果由(36.1)的 $\boldsymbol{v}_3$ 取 $(1.000, -0.200, 0.008)$，就得到

$$\lambda_R=6.4089,$$

大体上等于再进行2~3次迭代后得到的結果。

这一用 Rayleigh 商求特征值的近似值的方法，在振动問題和翹曲問題中常常使用。

2) **包含定理** (Einschliegungssatz) ❶　在(34.4)中代替 $\boldsymbol{x}$ 代入試解向量 $\boldsymbol{v}$，求

$$l_k=\frac{\boldsymbol{H}\boldsymbol{v}\ \text{的第}\ k\ \text{个分量}}{\boldsymbol{v}\ \text{的第}\ k\ \text{个分量}}\quad (k=1, 2, \cdots, n), \tag{39.3}$$

設 l_k 的(代数的)最大值和最小値分别为 l_{max}, l_{min}，那么，一定有满足

$$l_{min}\leqslant\lambda\leqslant l_{max} \tag{39.4}$$

❶ Collatz [9], p. 289.

的特征值 λ 存在。这就是所謂包含定理。如果 $\boldsymbol{v}$ 与某个特征向量 $\boldsymbol{x}_j$ 一致，那么所有的 l_k 都等于 λ_j。

这一定理在用松弛法把某几个 λ 的存在范圍縮小到某一程度时是有用的。

例 2 在松弛法中的应用。在 § 34 的例中，由(39.3)可以确定 l_k 如下：

$$\left.\begin{aligned} 6v_1-2v_2 \qquad\qquad &= l_1v_1, \\ -v_1+\frac{3}{2}v_2-\frac{1}{2}v_3 &= l_2v_2, \\ -\frac{1}{4}v_2+\frac{1}{4}v_3 &= l_3v_3. \end{aligned}\right\} \tag{39.5}$$

其中 $\boldsymbol{v}=(v_1, v_2, v_3)'$.

将試解向量的分量代入此式的左边进行計算，并分别用 $v_k(k=1, 2, 3)$ 除就得出此 l_k. 表 39.1 的上部表示出对应于 v_k 的增量 Δv_k 的 l_kv_k 的增量。例如，第 1 行表示，当 v_1 增加 1 时 l_1v_1 增加 6, l_2v_2 增加 -1, l_3v_3 增加 0. 下面的表是实际进行了松弛法的例子，最上边的一組对应于最小特征值 λ_1. 首先假定(20, 40, 100)′ 为試解向量 $\boldsymbol{v}$，而計算出了 l_kv_k 以及 l_k. 其次，粗体字(v_2 栏的 **50**)表示經过校正的分量。以此为基础，决定了这一列的 l_kv_k 乃至 l_k. 每次校正一个值。这样就得到

$$l_1=0.121, \quad l_2=0.129, \quad l_3=0.121, \tag{39.6}$$

由上述的包含定理可知，$0.121<\lambda_1<0.129$. 而經过改善后的試解向量为 $\boldsymbol{v}=(16.5, 48.5, 100)$. 就这个 $\boldsymbol{v}$ 求 λ_R 得到 $\lambda_R=0.1287165$，几乎成为精确解(参看例 1).

其次的表是关于 λ_2 的表，由类似的計算得到

$$1.21<\lambda_2<1.24, \quad \boldsymbol{v}=(42, 100, -26)'.$$

由此得到，$\lambda_R=1.212196$.

对于 λ_3 也可以作类似的計算。

3) 特征值的上下界定理 以下討論《变分法及其应用》§ 70 的应用。在上述的包含定理中，λ 的范圍显得过寬，因而以下試图将它縮小。λ_R 只給出了 λ 的上确界，在本定理中将給出它的下确界❶。

❶ 在本丛书《微分方程的近似解法》§ 24 中也給出了本定理的应用。

表 39.1

Δv_1	$\Delta(l_1v_1)$	Δv_2	$\Delta(l_2v_2)$	Δv_3	$\Delta(l_3v_3)$
1	6		−1		0
	−2	1	1.5		−0.25
	0		−0.5	1	0.25

v_1	l_1v_1	l_1	v_2	l_2v_2	l_2	v_3	l_3v_3	l_3
20	40	2	40	−10	−0.25	100	15	0.15
20	20	1	**50**	5	0.1	100	12.5	0.125
17	2	0.118	50	8	0.16	100	12.5	0.125
17	4	0.235	**49**	6.5	0.133	100	12.25	0.1225
16.5	1	0.061	49	7.0	0.143	100	12.25	0.1225
16.5	2	0.121	**48.5**	6.25	0.129	100	12.13	0.1213

$0.121 \leqslant \lambda \leqslant 0.129$;　(16.5, 48.5, 100)

v_1	l_1v_1	l_1	v_2	l_2v_2	l_2	v_3	l_3v_3	l_3
0	−200	$-\infty$	100	150	1.5	0	−25	$-\infty$
40	40	1.0	100	110	1.1	0	−25	$-\infty$
40	40	1.0	100	120	1.2	**−20**	−30	1.5
42	52	1.24	100	118	1.18	−20	−30	1.5
42	52	1.24	100	121	1.21	**−26**	−31.5	1.21

$1.21 \leqslant \lambda \leqslant 1.24$;　(42, 100, −26)

《变分法及其应用》§70 的(70.10)，(70.11)就是这个定理，用目前的記号写出就是

$$\left.\begin{aligned} &\lambda_R-\frac{\varepsilon^2}{\lambda_{j+1}-\lambda_R}\leqslant\lambda_j\leqslant\lambda_R+\frac{\varepsilon^2}{\lambda_R-\lambda_{j-1}},\\ &\varepsilon^2=\frac{(\boldsymbol{Hv})'\boldsymbol{B}(\boldsymbol{Hv})}{\boldsymbol{v}'\boldsymbol{Bv}}-\lambda_R^2. \end{aligned}\right\} \tag{39.7}$$

在此假定，特征值 λ_j 不是退化的❶，λ_R 和 λ_j 充分接近，而且 $\lambda_{j-1}<\lambda_R<\lambda_{j+1}$。如果 λ_j 的两个邻接的特征值 λ_{j-1}，λ_{j+1} 已知时，应用此式就可以把 λ_j 的存在范圍縮小，因此，与前述的松弛法計算結合起来就可以了。

❶ 特征值 λ_j 为特征方程的重根时称为是退化的 (degenerate)。——譯者注

例 3 由用松弛法求得的 $0.121\leqslant\lambda_1\leqslant0.129$，令 $\lambda_1\approx0.125$，用类似方法求得的 λ_3 满足 $6.4\leqslant\lambda_3\leqslant7.8$，由此取 $\lambda_3=7.1$。对应于 λ_2 的向量的近似值是 $\boldsymbol{v}=(0.42,\ 1.00,\ -0.26)$，由此得到

$$\lambda_R=1.212196,\quad \varepsilon^2=0.000051.$$

因此，由(39.7)得到

$$1.212146\leqslant\lambda_2\leqslant1.212205,$$

与最初的 $1.21\leqslant\lambda_2\leqslant1.29$ 比較这范圍已經很小了。

对于 $j=1$，即对于最小特征值，λ_R 本身就是它的上确界。

§40 圆柱形原子反应堆的临界計算

最后，作为特征值問題的例子，概略地叙述一下圆柱形原子反应堆的临界計算。这是用日本现在使用的具有最大容量的继电器計算机——电气試驗所的 ETL-Mark-II 計算出来的，这样也就可以看到，目前日本可以进行計算的程度❶。

計算对象的反应堆的尺碼如图 40.1 所示，其中的反射体是輕水。象这样的水减速原子反应堆，不是根据年龄扩散理論，而是根据 Selengut-Goertzel 方程的处理被认作是有效的。这样得到的方程是关于单位 lethargy 中中子束 φ 的扩散型方程，而 φ 是 lethargy u 和位置的函数。应用把能量群分为 4 群的分組理論 (multi-group method)❷加以变形，就得到如下的方程：

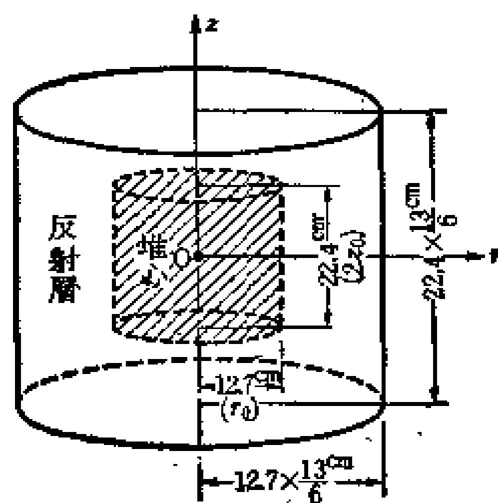


图 40.1

❶ 藤中惠，吉島重和：反射体付圆筒形原子炉的临界計算(带有反射层的圆柱形原子反应堆計算)，原子力发电，1, 4, pp. 12~18, (1957).

❷ 武谷，丰田：原子炉，p. 77；岩波讲座现代物理学(1955).

$$\left.\begin{aligned} &A_i\varphi_i - B_i\nabla^2\varphi_i = C_i\overline{\varphi}_{i-1} + z_i S, \\ &S = \nu\sum_{j=1}^{4} F_j\overline{\varphi}_j \quad (i=1,\ 2,\ 3,\ 4), \end{aligned}\right\} \tag{40.1}$$

其中 $\overline{\varphi}_i$ 是在能量级的分組区間 i 中在单位 lethargy 中的中子束的平均值，A_i, B_i, C_i 是由輸送截面积，吸收截面积，氫元素的散射截面积等决定的系数，z_i 是由裂变中子的能量分布决定的系数，F_j 是由裂变截面积和 lethargy 决定的系数，每个系数在堆心和反射层內都取不同的值。这些值可以由反应堆的各因素如表 40.1 所示那样計算出来。

表　40.1

i		1	2	3	4
堆心	A_i	0.2462	0.4662	1.454	0.1281
	B_i	5.985	2.967	7.838	0.1749
	C_i	0	0.2460	0.4660	1.321
	z_i	0.6997	0.3003	0	0
	F_i	0.0006005	0.0006058	0.08874	0.09291
反射层	A_i	0.2466	0.4679	1.348	0.01809
	B_i	6.132	3.013	7.969	0.1833
	C_i	0	0.2466	0.4678	1.331
	z_i	0	0	0	0
	F_i	0	0	0	0

ν 是在一次裂变中放出的中子平均数，按使用的燃料而取确定的值（在目前的情形 $\nu=2.46$），S 是表示由于裂变而产生的中子空間分布的发生項。又 $i=4$ 表示热中子的能量級。

$\overline{\varphi}_i$ 是处处有限的而且不能为負，此外还必需滿足下列边界条件：

a. 在反射层的外表面上，$\overline{\varphi}_i=0$,

b. 在堆心和反射层的界面上，

$$[\overline{\varphi}_i]_{\substack{r_0-0\\z_0-0}} = [\overline{\varphi}_i]_{\substack{r_0+0\\z_0+0}}, \quad [B_i\nabla\overline{\varphi}_i]_{\substack{r_0-0\\z_0-0}} = [B_i\nabla\overline{\varphi}_i]_{\substack{r_0+0\\z_0+0}}.$$

(式中 r_0，$2z_0$ 分别表示堆心的半徑和高。) 在上列边界条件下解(40.1)，只有在 ν 取某个特定的值 ν_c 时 $\bar{\varphi}_i$ 才有非零解。也就是說，ν_c 是这一系統的特征值，而对应于 ν_c 的 $\bar{\varphi}_i$ 是特征函数。只有在計算出的 ν_c 与 ν 一致时，反应堆才是临界的，而且存在有稳恒状态下的非零的有限中子束。当 $\nu>\nu_c$ 时由于供給的中子数超过了为使連鎖反应持續进行所必要的中子数因而成为发散的，当 $\nu<\nu_c$ 时，中子束逐步衰減，而在稳恒状态时变成了 0。ν/ν_c 表示使用了这种燃料时中子在每一代的增殖倍数，称为有效增殖倍数。

通常，临界計算的目的在于，关于給定的尺碼与燃料組成求出 ν_c 或 $\bar{\varphi}_i$，如果 ν_c 与 ν 沒有一致到所希望的范圍以內，那么就改变反应堆的尺碼和燃料組成而重新計算，这样反复下去一直到 ν_c 与 ν 一致到所希望的程度为止。在目前的情形，由于临界状态的反应堆的尺碼和燃料組成已經給定，因此，問題在于确定，計算出的 ν_c 以怎样的精确度与 ν 一致。

这問題的解法大体上可以进行如下：首先，作为第 0 近似假定出 S 的空間分布，那么，在(40.1)中令 $i=1$ 所得的方程成为关于

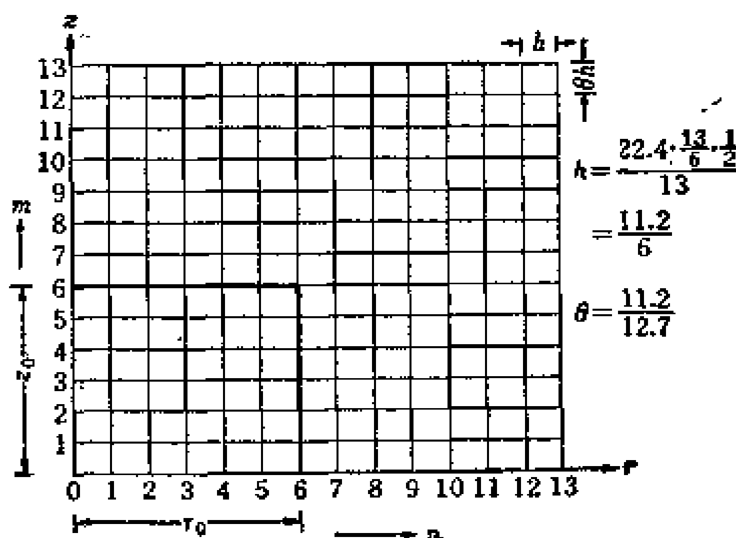


图 40.2

$\bar{\varphi}_1$ 的椭圓型方程，因而可以在給定的边界条件下求解。

由于目前問題的軸对称性，使用圓柱坐标，而用差分近似代換自变量 r 和 z 的 Laplace 算子。为此，考慮到对称性，将 r-z 坐标的二維区域如圖 40.2 那样分成 13×13 个网格，設 r 方向的步长为 h 而 z 方向的步长为 θh，那么，例如对于在一般网点 (m, n) 的 $\bar{\varphi}$，得到

$$\nabla^2\bar{\varphi}(m, n) = \frac{1}{h^2}\left\{\left(1+\frac{1}{2n}\right)\bar{\varphi}(n+1, m) + \left(1-\frac{1}{2n}\right)\bar{\varphi}(n-1, m) + \frac{1}{\theta^2}\bar{\varphi}(n, m+1) + \frac{1}{\theta^2}\bar{\varphi}(n, m-1) - \left(2+\frac{2}{\theta^2}\right)\bar{\varphi}(n, m)\right\}[1]. \tag{40.2}$$

通过这样的代换，就得到关于 $\bar{\varphi}_1$ 的 169 元联立方程。用 § 32 所述的 Liebmann 加快迭代法解这方程組，得到第一近似 $\bar{\varphi}_1^{(1)}$. 經过 4～6 次的迭代大体上就收斂了。每次迭代約需要 1 小时时間。

得到 $\bar{\varphi}_1^{(1)}$ 后，类似地可以求得 $\bar{\varphi}_2^{(1)}$, $\bar{\varphi}_3^{(1)}$, $\bar{\varphi}_4^{(1)}$. 每个都需要 4～5 次的迭代。由这样得到的 $\bar{\varphi}_i^{(1)}$ 可以計算 $\sum F_j\bar{\varphi}_j^{(1)}$. 如果这个值与最初假定的分布 $S^{(0)}$ 成比例，那么，$S^{(0)}$ 是正确的。比值 ν_c 是特征值，但一般說来不成比例。因此，例如可以重新假定分布 $S^{(1)}$ 如下：

$$S^{(1)} = \nu_c^{(0)}\sum_{j=1}^{4} F_j\bar{\varphi}_j^{(1)},\quad \nu_c^{(0)} = \int S_0 dv \Big/ \int [\sum F_j\bar{\varphi}_j^{(1)}] dv. \tag{40.3}$$

式中的积分是沿着堆心的容积积分，也应改写成差分形状而用数值方法积分。以这样得到的 $S^{(1)}$ 为基础順次求 $\bar{\varphi}_i^{(2)}$ $(i=1, 2, 3, 4)$，而計算 $\nu_c^{(1)}$ 和 $S^{(2)}$. 重复上述的过程一直到相邻接的两个 $S^{(k)}$

[1] 方程在对称軸与边界上的点的形状不同，关于这些問題詳見前引文献(藤中，吉島的論文，Shortley 的論文 (137 頁))，并参看 G. M. Roe: KAPL-950, Knolls Atomic Power Laboratory, 1954.

与 $S^{(k+1)}$ 的分布一致为止。此时 $\nu_c^{(k)}$ 成为特征值，而 $\varphi_i^{(k+1)}$ 給出中子束的分布。

实际上，由于計算时間的限制，关于 S 作了 4 次迭代就結束了。計算机械的实际工作时間約为 60 小时。

按照这样算得的結果，$S^{(3)}$ 和 $S^{(4)}$ 在堆心和反射层交界的棱部变动最大，約为 5%。变动少的地方在 1% 以下。这种状态大体上可以认为是滿意的。又 $\nu_c=2.496$，对于实驗所得的值 $\nu=2.46$，有效增殖倍数为 $\nu/\nu_c=0.9855$，与实驗值的誤差約为 1.5%。

把 10 进位的 10 位数，以及符号和表示小数位置的指数（小数点的移动可以到 ±19 位）构成的数作为一个"辞"，Mark-II 具有 200 个"辞"的記忆容量。在其他国家中，使用着与 Mark-II 不能相比的具有很大記忆容量以及計算速度的电子計算机，而且使用更多的分組数和网点数进行着本节所述的計算，因此，使用了在計算速度、容量、分組数、网点数等方面都受到很大限制的 Mark-II 計算机能进行如上的計算，而且得到相当精密的值，实在是值得重視的。

参 考 书

——在第1章至第4章中主要引用的参考书——

[1] P. S. Dwyer: Linear Computations (Wiley, 1951), 344.

[2] W. E. Milne: Numerical Calculus (Princeton University press, 1949), 393.

[3] F. B. Hildebrand: Introduction to Numerical Analysis (McGraw-Hill, 1956), 511.

[4] Z. Kopal: Numerical Analysis (Wiley, 1955), 556.

[5] A. D. Booth: Numerical Methods (Butterworths, 1955), 195.

[6] A. S. Householder: Principles of Numerical Analysis (McGraw-Hill, 1953), 274.

[7] D. R. Hartree: Numerical Analysis (Oxford Univ. Press, 1952), 287.

[8] K. S. Kunz: Numerical Analysis (McGraw-Hill, 1957), 381.

——在第5章至第7章中主要引用的参考书——

[9] L. Collatz: Eigenwertaufgaben mit Technischen Anwendungen (Akademische Verlagsgesell., 1949), 466.

[10] L. Collatz: Numerische Behandlung von Diffenrentialgleichungen (Springer Verlag, 1951), 458.

[11] S. H. Crandall: Engineering Analysis, A Survey of Numerical Procedures (McGraw-Hill, 1956), 417.

[12] F. B. Hildebrand: Methods of Applied Mathematics (Princeton-Hall, 1952), 523.

[13] R. A. Buckingham: Numerical Methods (Pitman, 1957), 597.

[14] W. E. Milne: Numerical Solution of Differential Equations (Wiley, 1953), 275.

[15] H. Levy and E. A. Baggot: Numerical Studies in Differential Equations (Watts), 238. (有 Dover 版及其日譯本、雨宫綾夫譯，河出书房，昭和 17＝1942.)

[16] 日高孝次：数値积分法(上,下),(岩波,昭和 11＝1936), 上 221. 下 292.

[17] 柴垣和三雄：常微分方程的数値解法(岩波,昭和 17＝1942), 162.

[18] 福原満洲雄等：常微分方程(本丛书)．
[19] 加藤敏夫等：微分方程的近似解法(本丛书)．
[20] 加藤敏夫：变分法及其应用(本丛书)．
[21] 森口繁一：穿孔卡計算机(本丛书)．
[22] C. Lanczos: Applied Analysis (Prentice Hall, 1956), 539. [特别是，載有关系于 § 13 (使最大誤差为最小的逼近)的便利的数值表(参看 § 13 的注)。]

校 后 記

張 鴻 游兆永

这本书介紹了数值計算法中一些常用的方法。在内容方面，若与已出版的国内編著的計算方法书籍相比較，一方面，有些部分（如积分方程、拟綫性双曲型方程解法等）沒有談到，有些部分（如方程近似解法等）談得过于簡略；但另一方面，則有些部分（如方法的計算程序等）讲得較仔細，有些部分（如数值表等）讲得較多些。例題（特别是数值例题）相当丰富，并注意到方法优缺点的分析。

在讲法方面，一般叙述簡练，而且看来作者不打算对所有問題都給出詳尽的理論分析与严密推演，所以有很多地方或是通过例題說明理論与方法；或是就特殊情况来証明理論与方法；或是在簡要地提出理論与方法之后再用例題来說明。我們感觉到，可能是作者有意識地为了使得数值計算实际工作者和初学者易于理解和便于应用而不多讲理論的。因此，就目前我国出版情况来說，这本书对于初学的讀者和进行数值計算实际工作的数学工作者及工程师是很有帮助的；而对于希望較为系統与深入地掌握数值計算法的理論与方法的讀者来說，也可以在閱讀其他理論与方法比較完备的书籍的同时把它作为一般的参考資料。

原书所列举的参考书中，西文书如[3]，[6]，[10]等都是很好的参考书（有些已有了俄譯本），又日文的书籍也列举了，但这些书大部分（特别是日文书）在国内不容易普遍看到，而且原书中也沒有提到中、俄文的参考书。所以下面再补充介紹几本常用的参考书，其中前一部分在一定程度上可以作为初学者的基本讀本（特别是[6]等的内容是非常丰富完备的）；后一部分可提供对計算方法

中某些方面問題有兴趣深入鉆研的讀者选讀。

(一)

[1] 中國科学院計算技术研究所：計算方法讲义(科学出版社，1958)。
[2] 胡祖熾：計算方法(高等教育出版社，1959)。
[3] 北大、吉大、南大計算数学教研室：計算方法(人民教育出版社，1961)。
[4] И. П. Мысовских: 計算方法(吉大計算数学教研室譯，人民教育出版社，1960)。
[5] 王德人等：計算实习(初等部分、高等部分)(高等教育出版社，1959)。
[6] И. С. Березин, Н. П. Жидков: Методы Вычислений, т. 1, 2(Физматгиз, 1959)。
[7] Б. П. Демидович, И. А. Марон: Основы Вычислительной Математики (Физматгиз, 1960)。
[8] Г. Н. Положего: Математический Практикум(Физматгиз, 1960)。
[9] Л. В. Канторович, В. И. Крылов: Приближенные Методы Высшего Анализа(Гостехиздат, 1952)。
[10] Anthony Ralston, Herbert S. Wilf: 数字計算机上用的数学方法(徐献瑜等譯，上海科学技术出版社，1963)。

(二)

[11] Д. К. Фаддеев, В. Н. Фаддеева: 綫代数計算方法 (殷国华等譯，上海科学技术出版社，即将出版)。
[12] E. Bodewiz: Matrix Calculus (North-Holland, 1956)。
[13] Ш. Е. Микеладзе: 数学分析的数値方法(科学出版社，1957)。
[14] В. К. Саульев: Интегрирование Уравнений Параболического Типа Методом Сеток(Физматгиз, 1960)。
[15] R. D. Richtmyer: Difference Methods For Initial-Value Problems (Interscience, 1957, 有俄譯本)。
[16] В. С. Рябенький, А. Ф. Филиппов: Об Устойчивости Разностных Уравнений(гостехиздат, 1956)。
[17] Н. П. Бусленко, Ю. А. Шрейдер: 蒙德卡洛方法(王毓云等譯，上海科学技术出版社，即将出版)。